The American Economic Review

PAPERS AND PROCEEDINGS

OF THE

Eighty-first Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

Chicago, Illinois, December 28-30, 1968

CALCUTTA UNIVERSITY CENTRAL LIBRARY

Edited by the Secretary of the Association

MAY 1969

60

391
501

AMERICAN ECONOMIC ASSOCIATION

Organized at Saratoga, New York, September 9, 1885

PAST OFFICERS

Presidents

1. Francis A. Walker,* M.I.T., 1886-92
2. Charles F. Dunbar,* Harvard, 1893
3. John B. Clark,* Columbia, 1894-95
4. Henry C. Adams,* Michigan, 1896-97
5. Arthur T. Hadley,* Yale, 1898-99
6. Richard T. Ely,* Wisconsin, 1900-01
7. Edwin R. A. Seligman,* Columbia, 1902-03
8. Frank W. Taussig,* Harvard, 1904-05
9. Jeremiah W. Jenks,* Cornell, 1906-07
10. Simon N. Patten,* Pennsylvania, 1908
11. Davis R. Dewey,* M.I.T., 1909
12. Edmund J. James,* Illinois, 1910
13. Henry W. Farnam,* Yale, 1911
14. Frank A. Fetter,* Princeton, 1912
15. David Kinley,* Illinois, 1913
16. John H. Gray,* Minnesota, 1914
17. Walter F. Willcox,* Cornell, 1915
18. Thomas N. Carver,* Harvard, 1916
19. John R. Commons,* Wisconsin, 1917
20. Irving Fisher,* Yale, 1918
21. Henry B. Gardner,* Brown, 1919
22. Herbert J. Davenport,* Cornell, 1920
23. Jacob H. Hollander,* Johns Hopkins, 1921
24. Henry R. Seager,* Columbia, 1922
25. Carl C. Plehn,* California, 1923
26. Wesley C. Mitchell,* Columbia, 1924
27. Allyn A. Young,* Harvard, 1925
28. Edwin W. Kemmerer,* Princeton, 1926
29. Thomas S. Adams,* Yale, 1927
30. Fred M. Taylor,* Michigan, 1928
31. Edwin F. Gay,* Harvard, 1929
32. Matthew B. Hammond,* Ohio State, 1930
33. Ernest L. Bogart,* Illinois, 1931
34. George E. Barnett,* Johns Hopkins, 1932
35. William Z. Ripley,* Harvard, 1933
36. Harry A. Millis,* Chicago, 1934
37. John M. Clark,* Columbia, 1935
38. Alvin S. Johnson, New School, 1936
39. Oliver M. W. Sprague,* Harvard, 1937
40. Alvin H. Hansen, Harvard, 1938
41. Jacob Viner, Chicago, 1939
42. Frederick C. Mills,* Columbia, 1940
43. Sumner H. Slichter,* Harvard, 1941
44. Edwin G. Nourse, Brookings, 1942
45. Albert B. Wolfe,* Ohio State, 1943
46. Joseph S. Davis, Stanford, 1944
47. I. L. Sharfman, Michigan, 1945
48. E. A. Goldenweiser,* Institute for Advanced Study, 1946
49. Paul H. Douglas, Chicago, 1947
50. Joseph A. Schumpeter,* Harvard, 1948
51. Howard S. Ellis, California, 1949
52. Frank H. Knight, Chicago, 1950
53. John H. Williams, Harvard, 1951
54. Harold A. Innis,* Toronto, 1952
55. Calvin B. Hoover, Duke, 1953
56. Simon Kuznets, Pennsylvania, 1954
57. John D. Black,* Harvard, 1955
58. Edwin E. Witte,* Wisconsin, 1956
59. Morris A. Copeland, Cornell, 1957
60. George W. Stocking, Vanderbilt, 1958
61. Arthur F. Burns, Columbia, 1959
62. Theodore W. Schultz, Chicago, 1960
63. Paul A. Samuelson, M.I.T., 1961
64. Edward S. Mason, Harvard, 1962
65. Gottfried Haberler, Harvard, 1963
66. George J. Stigler, Chicago, 1964
67. Joseph J. Spengler, Duke, 1965
68. Fritz Machlup, Princeton, 1966
69. Milton Friedman, Chicago, 1967
70. Kenneth E. Boulding, Colorado, 1968

Secretaries

Richard T. Ely,* 1886-92
Edward A. Ross,* 1893
Jeremiah W. Jenks,* 1894-96
Walter F. Willcox,* 1897-99

Treasurers

Edwin R. A. Seligman,* 1886-90
Frederick B. Hawley,* 1891-95
Charles H. Hull,* 1896-99

Secretary-Treasurers

Charles H. Hull,* 1900
Frank A. Fetter,* 1901-06
Winthrop M. Daniels,* 1907-08
Thomas N. Carver,* 1909-13
Allyn A. Young,* 1914-20
Ray B. Westerfield,* 1921-25
Frederick S. Deibler,* 1925-35
James Washington Bell,* 1936-61

Editors

Davis R. Dewey,* 1911-40
Paul T. Homan, 1941-51
Bernard F. Haley, 1952-62
John G. Gurley, 1962-68

* Deceased.

OFFICERS FOR 1969

President

William J. Fellner, Yale University

President-elect

Wassily Leontief, Harvard University

Vice-Presidents

Moses Abramovitz, Stanford University
Joe S. Bain, University of California, Berkeley

Secretary-Treasurer and Editor of Proceedings

Harold F. Williamson, Northwestern University

Managing Editor of the American Economic Review

George H. Borts, Brown University

Editor of the Journal of Economic Literature

Mark Perlman, University of Pittsburgh

Elected Members of the Executive Committee

Kenneth J. Arrow, Harvard University

Franco Modigliani, Massachusetts Institute of Technology

Otto Eckstein, Harvard University

Anthony Scott, University of British Columbia

Mary Jean Bowman, University of Chicago

Charles L. Schultze, Brookings Institution

Ex Officio Members

Milton Friedman, University of Chicago
Kenneth E. Boulding, University of Colorado

Counsel

Seymour J. Rubin, Washington, D.C.

CUK-H03124-60-P270922

THE AMERICAN
ECONOMIC REVIEW

OL. LIX MAY, 1969 NUMBER 2

PAPERS AND PROCEEDINGS

OF THE

Eighty-first Annual Meeting

60

OF THE

AMERICAN ECONOMIC ASSOCIATION

Chicago, Illinois, December 28–30, 1968

Edited by HAROLD F. WILLIAMSON, *Secretary of the Association*
and
GERTRUDE TAIT, *Executive Assistant*

Copyright American Economic Association, 1969

PRINTED BY GEORGE BANTA COMPANY, INC.

Publication Office: Curtis Reed Plaza, Menasha, Wisconsin 54952

Executive Office: American Economic Association
629 Noyes Street, Evanston, Illinois 60201

Inquiries and other communications regarding membership, meetings and the general affairs of the Association, as well as orders for publications, should be addressed to Harold F. Williamson, Secretary of the American Economic Association, 629 Noyes Street, Evanston, Illinois 60201.

Changes of address should be sent to the Secretary at the Executive Office at 629 Noyes Street, Evanston, Illinois. To be effective, notice must reach the office by the 1st of the month previous to the month of publication. The Association cannot be responsible for copies lost due to failure to report change of address in time.

Entered at the post office at Menasha, Wisconsin, as second class matter. Acceptance for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (D) provided for in the Act of February 28, 1925, embodied in paragraph 4, section 412, P. L. and R., authorized September 13, 1928.

The *American Economic Review* is published five times a year, in March, May, June, September, and December, and is sent to all members of the American Economic Association as one of the privileges of membership, $8.00 of the annual membership dues of $10.00 being in payment of a year's subscription to the publication. Printed in U.S.A.

PRICE $4.00

TABLE OF CONTENTS

Program of the Eighty-first Annual Meeting vii

PAPERS

Richard T. Ely Lecture

History of Economic Doctrines and Economic History *Alexander Gerschenkron* 1

Theory of Innovation

An Economic Theory of Technological Change *William D. Nordhaus* 18
Classificatory Notes on the Production and Transmission of Technological Knowledge *Kenneth J. Arrow* 29
Technological Progress and Microeconomic Theory *Werner Z. Hirsch* 36
Discussion
Evsey D. Domar 44
Joseph Stiglitz 46

Research and Development and Other Determinants of Investment

Investment and the Frustrations of Econometricians *Robert Eisner* .. 50
Industrial Research and Development: Characteristics, Costs, and Diffusion of Results *Edwin Mansfield* 65
Market Structure and the Stability of Investment *Frederic M. Scherer* 72
Research and Development, Production Functions, and Rates of Return *Jora R. Minasian* 80
Discussion
Jora R. Minasian 86

Theory of the Firm and of Market Structures

Advertising and the Advantages of Size *William S. Comanor* and *Thomas A. Wilson* 87
Survival-Ability as a Test of Efficiency *Joe S. Bain* 99
Allocative Efficiency and the Limits of Antitrust *Oliver E. Williamson* 105
Discussion
Richard B. Heflebower 119
Lester G. Telser 121

Wage-Price Dynamics, Inflation, and Unemployment

The Phillips Curve and the Distribution of Unemployment *G. C. Archibald* 124
Improving the Labor Market Trade-off between Inflation and Unemployment *Charles C. Holt* 135
The New Microeconomics in Inflation and Employment Theory *Edmund S. Phelps* 147
Discussion
Frank Brechling 161
Otto Eckstein 162
Richard A. Lester 164

The Supply Function in Agriculture Revisited

The Representative Farm Approach to Estimation of Supply Response *Jerry A. Sharples* 168
Positivistic Measures of Aggregate Supply Elasticities: Some New Approaches *Luther G. Tweeten* and *C. Leroy Quance* 175
Discussion
Marvin Kottke 184
W. Neill Schaller 185
James H. White 187

Problems in the Theory of Public Choice

Social Cost and Government Action *Gordon Tullock* 189
An Economic Theory of Social Movements *Albert Breton* and *Raymond Breton* 198
Lewis Carroll and the Theory of Games *Duncan Black* 206
Discussion
Thomas E. Borcherding 211
A. Allan Schmid 212
Benjamin Ward 214

The Efficiency of Education in Economics

The Efficiency of Programmed Learning in Teaching Economics: The Results of a Nationwide Experiment *Richard E. Attiyeh*, *G. L. Bach*, and *Keith G. Lumsden*. 217
Performance on the New Test of Understanding in College Economics *Arthur L. Welsh* and *Rendigs Fels* 224
DEEP: Strengthening Economics in the Schools *John E. Maher* 230
Discussion
Lloyd G. Reynolds 239
Peter O. Steiner 240

Carey C. Thompson 242

Transportation and the Public Utilities

Transportation Regulation and Economic Efficiency *Martin T. Farris* 244
Congestion Theory and Transport Investment *William S. Vickrey.* 251
Transportation and Price Stability *George W. Wilson* 261
Discussion
Donald V. Harper 270
John L. Hazard 271
Joseph V. Yance 274

Monetary Theory

The Lag in Monetary Policy as Implied by the Time Pattern of Monetary Effects on Interest Rates *Phillip Cagan* and *Arthur Gandolfi* 277
Money and Permanent Income: Some Empirical Tests *James Tobin* and *Craig Swan* 285
Econometric Analysis of Stabilization Policies *Albert Ando* and *Franco Modigliani* 296
Discussion
Richard G. Davis 315
David Meiselman 318
Donald P. Tucker 321

The Future of Gold

Real Gold, Dollars, and Paper Gold *Robert A. Mundell* 324
Speculation on Gold Speculation *Fritz Machlup* 332
The Gold Rush of 1968 in Retrospect and Prospect *Harry G. Johnson.* 344
Discussion
Edward M. Bernstein 349
Fred Hirsch 351
Miroslav A. Kriz 353

Round Table on Exchange Rate Policy

Gottfried Haberler 357
Henry C. Wallich 360
Peter B. Kenen 362
Milton Friedman 364
Fritz Machlup 366

Economic History

The Cost to America of British Imperial Policy *Peter D. McClelland* 370
Discussion
Jonathan R. T. Hughes 382
Herman E. Krooss 384

Economic Development

Economic Development in Historical Perspective *John C. H. Fei* and *Gustav Ranis* 386
The Content of Development Economics *Lloyd G. Reynolds* 401
Instruments and Goals in Economic Development *Irma Adelman*, *Marsha Geier*, and *Cynthia Taft Morris* 409
Discussion
Robert E. Baldwin 427
Charles R. Frank, Jr 429
Carlos F. Diaz-Alejandro 432

Problems in the Area of Poverty

The OBE Size Distribution Series: Methods and Tentative Results for 1964 *Edward C. Budd* and *Daniel B. Radner* 435
Measures of Economic Well-offness and Their Correlates *James N. Morgan* and *James D. Smith* 450
Graduated Work Incentives: An Experiment in Negative Taxation *Harold W. Watts* 463
Discussion
Christopher Green 473
Harold W. Guthrie 474

Strategic Theory and Its Applications

The Principle of "Fiscal Equivalence": The Division of Responsibilities among Different Levels of Government *Mancur Olson, Jr.* 479
Models of Segregation *Thomas C. Schelling* 488
Strategic Interaction and Resource Allocation in Metropolitan Intergovernmental Relations *Jerome Rothenberg* 495

Round Table on Allocation of Resources in Law Enforcement

Anthony Downs 504
William M. Landes 505
Simon Rottenberg 508
Richard B. Hoffman 510

Centralization and Decentralization in Economic Systems

On the Concept and Possibility of Informational Decentralization *Leonid Hurwicz* 513

On the Comparison of Centralized and Decentralized Economies *Thomas Marschak* 525
Discussion
Bela Balassa 533
Abram Bergson 537

Completed Dissertation Research

Manpower Shortages in Local Government Employment *Eugene J. Devine* 538
Resource Allocation with Probabilistic Individual Preferences *Richard Zeckhauser* 546
Effects of Taxation on Risk-Taking *Aiko N. Shibata* 553
Discussion
David Cass 562
William H. Oakland 563
Mark V. Pauly 565

PROCEEDINGS

Minutes of the Business Meeting **571**

Minutes of the Executive Committee Meetings **574**

Reports

Secretary *Harold F. Williamson* ... 578
Treasurer *Harold F. Williamson* .. 584
Committee on Financial and Investment Policies *Milton Friedman* .. 586
Finance Committee *C. Wells Farnham* 587
Auditor *Arthur Andersen & Co.* ... 589
Managing Editor, *American Economic Review* *John G. Gurley* ... 594
Managing Editor, *Journal of Economic Abstracts* *Arthur Smithies* 599
Committee on Economic Education *G. L. Bach* 600
Representative to the National Bureau of Economic Research *Willard L. Thorp* 602
Representative to the International Economic Association *Fritz Machlup* 604
Economics Institute Policy and Advisory Board *Henry Rosovsky* 605
Visiting Scientist Program in Economics *Phillip Saunders* 606
Census Advisory Committee *Solomon Fabricant* 607
Representative to the National Research Council, Division of Behavioral Sciences *James N. Morgan* 608

Publications of the American Economic Association **611**

PROGRAM OF THE EIGHTY-FIRST ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

Chicago, Illinois, December 28–30, 1968

Friday, December 27, 1968
12:30 P.M.
Executive Committee Meeting

Saturday, December 28, 1968
8:30 A.M.
The Supply Function in Agriculture, Revisited (Joint session with the American Farm Economic Association)
Chairman: Glenn L. Johnson, Michigan State University
Papers: Jerry A. Sharples, U.S. Department of Agriculture; Luther G. Tweeten and Leroy Quance, Oklahoma State University
Discussants: Marvin W. Kottke, University of Connecticut; W. Neill Schaller, U.S. Department of Agriculture; James H. White, University of Arkansas
Transportation and the Public Utilities (Joint session with the Transportation and Utilities Group)
Chairman: Charles A. Taff, University of Maryland
Papers: Martin T. Farris, Arizona State University; William S. Vickrey, Columbia University; George W. Wilson, Indiana University
Discussants: Donald V. Harper, University of Minnesota; John L. Hazard, Michigan State University; Joseph V. Yance, Boston University

10:30 A.M.
Wage-Price Dynamics, Inflation, and Unemployment
Chairman: Almarin Phillips, University of Pennsylvania
Papers: G. C. Archibald, University of Essex; Charles C. Holt, University of Wisconsin; Edmund S. Phelps, University of Pennsylvania
Discussants: Frank Brechling, Northwestern University; Otto Eckstein, Harvard University; Richard A. Lester, Princeton University
Problems in the Theory of Public Choice (Joint session with the Public Choice Society)
Chairman: James M. Buchanan, University of California, Los Angeles
Papers: Gordon Tullock, Virginia Polytechnic Institute; Albert Breton, London School of Economics, and Raymond Breton, Johns Hopkins University; Duncan Black, University of Wales and University of Chicago
Discussant: Thomas Borcherding, University of Washington; Allan Schmid, Michigan State University; Benjamin Ward, University of California, Berkeley
Microdevelopment: The Transformation of Village-level Economies (Joint session with the Association for Comparative Economics)
Chairman: George Dalton, Northwestern University
Papers: Clifton R. Wharton, Jr., Agricultural Development Council; Barney Schwalberg, Brandeis University
Discussant: John Q. Adams, University of Maryland

2:30 P.M.
Theory of Innovation (Joint session with the Econometric Society)
Chairman: Edmund S. Phelps, University of Pennsylvania
Papers: William D. Nordhaus, Yale University; Kenneth J. Arrow, Harvard University; Werner Z. Hirsch, University of California, Los Angeles
Discussants: Evsey D. Domar, Massachusetts Institute of Technology; Joseph E. Stiglitz, Yale University
Economic Development
Chairman: Arnold C. Harberger, University of Chicago
Papers: John C. H. Fei, Cornell University, and Gustav Ranis, Yale University; Lloyd G. Reynolds, Yale University; Irma Adelman, Northwestern University, Marsha Geier, Yale University, and Cynthia T. Morris, American University
Discussants: Robert E. Baldwin, University of Wisconsin; Charles R. Frank, Princeton University; Carlos F. Diaz-Alejandro, University of Minnesota
The Efficiency of Education in Economics
Chairman: Robert A. Gordon, University of California, Berkeley
Papers: Richard Attiyeh, University of California, San Diego, G. L. Bach and Keith G. Lumsden, Stanford University; Rendigs Fels, Vanderbilt University, and Arthur Welsh, Joint Council on Economic Education; John E. Maher, Joint Council on Economic Education and State University of New York
Discussants: Lloyd G. Reynolds, Yale University; Peter O. Steiner, University of Wisconsin; Carey C. Thompson, University of Texas

8:00 P.M.

Richard T. Ely Lecture

Chairman: Wassily Leontief, Harvard University

Paper: Alexander Gerschenkron, Harvard University

Sunday, December 29, 1968

8:30 A.M.

Problems in the Area of Poverty

Chairman: Mancur Olson, Jr., Department of Health, Education, and Welfare

Papers: Edward C. Budd, Pennsylvania State University, and Daniel B. Radner, Office of Business Economics; James N. Morgan, University of Michigan, and James D. Smith, Pennsylvania State University and Office of Economic Opportunity; Harold W. Watts, University of Wisconsin

Discussants: Christopher Green, North Carolina State University; Harold W. Guthrie, University of Illinois; Lester C. Thurow, Massachusetts Institute of Technology

Selected Graduate Student Papers (Joint session with Omicron Delta Epsilon)

Chairman: John D. Guilfoil, New York University

Papers: John Pisciotta, University of Texas; Raveendr Nath Batra, Southern Illinois University; Matthew Drennan, New York University

10:30 A.M.

Monetary Theory

Chairman: John G. Gurley, Stanford University

Papers: Phillip D. Cagen and Arthur Gandolfi, Columbia University; James Tobin and Craig Swan, Yale University; Albert K. Ando, University of Pennsylvania, and Franco Modigliani, Massachusetts Institute of Technology

Discussants: Richard G. Davis, Federal Reserve Bank of New York; David Meiselman, Macalaster College; Donald P. Tucker, University of Chicago

Theory of the Firm and of Market Structures

Chairman: Anthony D. Scott, University of British Columbia

Papers: William S. Comanor, Harvard University, and Thomas A. Wilson, University of Toronto; Joe S. Bain, University of California, Berkeley; Oliver E. Williamson, University of Pennsylvania

Discussants: Richard B. Heflebower, Northwestern University; Lester G. Telser, University of Chicago

Economic History

Chairman: Alexander Gerschenkron, Harvard University

Papers: Peter D. McClelland, Harvard University; Robert W. Fogel,[1] University of Chicago

Discussants: Jonathan R. T. Hughes, Northwestern University; Herman E. Krooss, New York University

12:30 P.M.

Joint Luncheon with the American Finance Association

Chairman: Walter E. Hoadley, Vice-President, A.F.A.

Presiding: Harry C. Sauvain, President, A.F.A., and Kenneth E. Boulding, President, A.E.A.

Speaker: Arjay Miller, Ford Motor Company

2:00 P.M.

Research and Development and Other Determinants of Investment (Joint session with the Econometric Society)

Chairman: Jora R. Minasian, University of Southern California

Papers: Robert Eisner, Northwestern University; Edwin Mansfield, University of Pennsylvania; Frederic M. Scherer, University of Michigan; Jora R. Minasian, University of Southern California

Discussants: Jora R. Minasian, University of Southern California

2:30 P.M.

Strategic Theory and Its Applications

Chairman: James M. Buchanan, University of California, Los Angeles

Papers: Mancur Olson, Jr., Department of Health, Education, and Welfare; Thomas C. Schelling, Harvard University; Jerome Rothenberg, Massachusetts Institute of Technology

Discussants: Joseph L. Bower,[1] Harvard University; Daniel Ellsberg,[1] RAND Corporation

Round Table on Exchange Rate Policy

Chairman: Fritz Machlup, Princeton University

Discussants: Gottfried Haberler, Harvard University; Henry C. Wallich, Yale University; PETER B. KENEN, Columbia University; MILTON FRIEDMAN, University of Chicago; FRITZ MACHLUP, Princeton University

8:00 P.M.
Presidential Address[2]
Chairman: ALBERT G. HART, Columbia University
Presidential Address: KENNETH E. BOULDING, University of Colorado

9:15 P.M.
Business Meeting

Monday, December 30, 1968
9:00 A.M.
Centralization and Decentralization in Economic Systems
Chairman: RICHARD W. JUDY, University of Toronto
Papers: LEONID HURWICZ, University of Minnesota; THOMAS A. MARSCHAK, University of California, Berkeley
Discussants: BELA BALASA, Johns Hopkins University; ABRAM BERGSON, Harvard University
Completed Dissertation Research
Chairman: WILLIAM R. ALLEN, University of California, Los Angeles
Papers: EUGENE J. DEVINE, Simon Fraser University; RICHARD ZECKHAUSER, Harvard University; AIKO N. SHIBATA, University of York
Discussants: DAVID CASS, Yale University; WILLIAM OAKLAND, Johns Hopkins University; MARK V. PAULY, Northwestern University
11:00 A.M.
Round Table on the Allocation of Resources in Law Enforcement
Chairman: THOMAS C. SCHELLING, Harvard University
Discussants: ANTHONY DOWNS, JR., Real Estate Research Corporation; WILLIAM M. LANDES, University of Chicago; SIMON ROTTENBERG, American Bar Foundation and Duke University; RICHARD B. HOFFMAN, State University of New York, Buffalo
The Future of Gold
Chairman: GOTTFRIED HABERLER, Harvard University
Papers: ROBERT A. MUNDELL, University of Chicago; FRITZ MACHLUP, Princeton University; HARRY G. JOHNSON, London School of Economics and University of Chicago
Discussants: EDWARD M. BERNSTEIN, EMB Ltd.; FRED HIRSCH, International Monetary Fund; MIROSLAV A. KRIZ, First National City Bank

---

[1] No manuscript received.
[2] Published in the March, 1969, *A.E.R.*

THE purpose of the American Economic Association, according to its charter, is the encouragement of economic research, the issue of publications on economic subjects, and the encouragement of perfect freedom of economic discussion. The Association as such takes no partisan attitude, nor does it commit its members to any position on practical economic questions. It is the organ of no party, sect, or institution. Persons of all shades of economic opinion are found among its members, and widely different issues are given a hearing in its annual meetings and through its publications. The Association, therefore, assumes no responsibility for the opinions expressed by those who participate in its meetings. Needless to say, the papers presented are the personal opinions of the authors and do not commit the organizations or institutions with which they are associated.

HAROLD F. WILLIAMSON
Secretary

CALCUTTA UNIVERSITY LIBRARY

RICHARD T. ELY LECTURE

HISTORY OF ECONOMIC DOCTRINES AND ECONOMIC HISTORY

Quantula theoria economica mundus regebatur.

By ALEXANDER GERSCHENKRON
Harvard University

Since the title of my lecture is not quite self-explanatory, I must first say a few words about the nature of my subject. Ideally, a presentation of this kind should emerge from the lecturer's current research work. Unfortunately, however, the problems of constructing an index of Russian industrial output from 1762 to 1914 (which happen to be my present concern), fascinating as they appear to me, are of negligible interest to the profession. I have decided, therefore, to have recourse to a general subject that still relates to my basic preoccupation with the flow of modern (essentially nineteenth century) European economic history, but connects it with the history of economic doctrines. The decision has been both a painful and an uncertain one. The intricacies of the Russian index would have bored and depressed you; the superficiality of what I shall try to say in all probability will disgust you, and, for all I know, may bore you only slightly less. Thus, the only indubitable benefit from this evening that I can promise you is a full revelation of the differential interpretations of the Harvard accent as displayed by Professor Leontief and myself.

Let me begin by saying that in our time the history of economic doctrines has come upon evil days. To some extent, this may be a still lingering result of the Keynesian revolution. Revolutions in general are unkind to history. It takes them a long time before they discover that they themselves are but another thread in the seamless web. More important, probably, has been the influence of science, and primarily of physics. The Department of Physics at Harvard has completely eliminated history of physics from its curriculum; such history has been shifted to an independent History of Science Department. By contrast, in the Department of Political Science, history of political thought is still the daily bread of the discipline. Today's economics finds itself between those extremes, but certainly not in the middle. We are getting closer and closer to physics. I find this neglect of doctrinal history unfortunate for a number of reasons. From a pedagogical point of view, I feel, translation into modern terms of past contributions to economics is a very effective way to teach modern economics. Beyond that, history of economic analysis is not simply a history of intellectual progress. It is also a history of shifts of emphasis and interest, of to-ing and fro-ing, of abandoned and rediscovered problems, of parallel and disparate strains. As such it may well be a continual source of inspiration for new departures and directions in economic theory.

As I set out to write this paper, I believed that demonstrating the strong impact of economic doctrines upon economic history could serve as an additional argument for greater preoccupation with the history of doctrines. Unfortunately, the paper as it has emerged fails to justify this expectation. The reason, I think, is twofold. On the one hand, it lies in the nature of some of the past economic doctrines; on the other hand, and more importantly, it must be sought in the gap that separated economic analysis from economic history and was filled by ideology, at least in one sense of that confusing but indispensable term. I am referring to ideology in the sense of choices for action (or inaction), with the vexed problems of distortion and perversion, intellectual inertia and intellectual fraud grafted upon it —problems, that is, which preoccupied many thinkers from Francis Bacon to Mannheim and beyond. For it is mainly as ideology as varying mixtures of theories, or rather paratheories, conjoined with many other ingredients that economic doctrines may have had practical significance by hastening or obstructing, as the case may be, the processes of economic change, the subject matter of economic history. Ideology then does two things: It reveals the impact of a doctrine, but also disguises its perversion, if not abandonment.

This is a topic of immense scope and complexity, and I am aware that I am broadening it even more when I say that I should like also to say something about history of economic doctrines from the point of view of study of economic history. All I can possibly do is to touch on a few aspects of the problem and offer some disjointed assertions and episodic illustrations. Fragmentary treatment and avowedly one-sided emphasis cannot yield more than highly tentative conclusions. In what follows, therefore, my thesis regarding the limited role of economic analysis in influencing economic history can indeed be suggested, but it cannot be established in a manner likely to satisfy our sense of reasoned adequacy.

One more remark to prevent possible misunderstandings. Nothing I shall say shall be taken to deny the usefulness, in fact, the necessity of, applying the tools of economic analysis to the study of economic history. On the contrary, I am convinced that such an application, known now as New Economic History, is far and away the best thing that has happened to the discipline for a long time. Nor am I interested here in the way in which economic conditions may have influenced the progress of economic analysis.[1] What I am concerned with here relates entirely to the impact of past economic doctrines on past events and also with certain dangers that arise when economic historians too readily regard contemporaneous economic theories as an integral part of the economic history of the times.

One of the difficulties an economic historian experiences with the history of economic thought is precisely that at times, at least, it serves to obscure rather than clarify the course of events and its meaning. I believe that mercantilism is a case in point, and that regarding mercantilism from an analytical point of view has tended to make it rather difficult to perceive its role and function in economic history. And it does not make much difference whether the attitudes to mercantilist theorizing were negative or more positive. In the hands of Jacob Viner, who continued, or returned to, a tradition dating at least since 1776, mercantilism appeared as a variety of fallacies, a remarkable chapter in the history of human stupidity, as something done, as Erasmus used to say, *a stultis apud stultos*, by fools for fools. Keynes, on the other hand, believed that in the peculiar circumstances of the time, interest in the influx of precious metals was quite sensible. And, after Keynes, Schumpeter in his more detailed investigations, came to the conclusion that there was not, or rather need not have been, any spectacular break from "the embryonic fund of theoretical knowledge" to the theories of classical economists.[2] My point here is not that Viner, or

---

[1] This problem is the subject matter of the recent symposium edited by Robert V. Eagly, *Events, Ideology, and Economic Theory* (Detroit, 1968).

[2] Jacob Viner, *Studies in the Theory of International Trade* (New York, 1937), Chaps. I and II; Joseph A. Schumpeter, *History of Economic Analysis* (New York, 1954), p. 376.

Keynes, or Schumpeter were right or wrong in their evaluations of mercantilist thought. This is not the issue at all. The point rather is that concentrating on the theoretical views of mercantilists—that is, essentially problems of foreign trade and money—does not open up the proper avenue for appraising mercantilism as a significant sequence of events in the economic history of Europe, particularly if it is desired to see the phenomenon in diversifed unity from the shores of the Atlantic to the plains of Russia.

By contrast, economic historians—and above all Heckscher—did lay the basis for a much broader and more effective interpretation. By stressing the primacy of the political factor, not only some aspects of mercantilist theorizing became more comprehensible, but the weight of emphasis was shifted from foreign economic policies to the area of domestic economic problems. Even Heckscher's interpretation, however, contained some inadequacies in the light of his stated intention to make "a contribution to the history of economic policy as *a common European problem*."[3] Because in reality Heckscher did not deal with the whole of Europe and concentrated essentially on the west of the continent, he was forced time and again to weaken his stress on power in favor of opulence as a parallel aim of mercantilist policies. Viewing mercantilism as a common European phenomenon would justify a different interpretation, and because of its bearing on the significance of mercantilistic theories, I should like to summarize my views very briefly, confining myself to three points.

1. I believe it is correct to say that as you move eastward across Europe, that is to say, along a line of increasing economic backwardness, you find a steadily increasing importance of power and a diminishing importance of wealth as motives and aims of governmental policies. As a result, as you reach Russia and study her momentous mercantilistic experience in the early eighteenth century, power can be seen as the exclusively dominating factor in shaping economic policies and, by the same token, in determining the nature and the results of economic change.

2. While recognizing the inconsistencies and ineptitudes of mercantilistic economic policies, some of the former having often been the result of conflicting time horizons (e.g., with respect to choices as between investment and military expenditures or between pressing fiscal needs and abolition of internal duties), I feel those policies must be seen as primarily oriented to economic development. It is, however, a curious peculiarity of those policies that while promoting economic growth, they were at the same time creating specific obstacles to subsequent economic development. I believe that the strength of these obstacles, too, can be related to the degree of economic backwardness of the countries concerned. In Russia, the mercantilistic experiment left the country with the consolidated, expanded, and effectively policed serfdom of the peasantry which had been one of the carrying pillars of mercantilistic policies but was to become a formidable hindrance to modern economic growth. Austrian mercantilism under Joseph II greatly increased the power of the central state and its role in economic development, but it was under Joseph's successors in the early nineteenth century that the centralized state became a main obstacle to railroad building—an obstacle, one must add, however, very much weaker than the serfdom of the peasantry in Russia, but at the same time very much stronger than the comparable survivals or aftereffects of mercantilistic policies in France or England. In fact, it can be argued, paradoxical as it may

---

[3] Eli F. Heckscher, *Mercantilism* (London, 1935), Vol. I, p. 13.

seem, that it was the very weakness of the obstacles left over by mercantilistic policies in more advanced countries that made possible Adam Smith's destructive attack upon them.

3. It seems to me that a further generalization can be hazarded that the more backward a country, the less important was the concern in its mercantilistic policies with any foreign economic policies and, by the same token, with attempts at analytical explanation, direction, or justification of these policies. In Russia, foreign economic policies hardly played any role until the very end of Peter the Great's reign, and the interest in mercantilistic literature was virtually nil under this most vigorous of all mercantilistic statesmen. Mercantilism *sans doctrine* was an indubitable historical fact.

Thus, my final conclusion is twofold. If an economic historian desires to obtain a unified view of mercantilism as a "common European phenomenon," it is the power of the state and policies of economic development pursued for the purpose of augmentation of power that he must regard as the common denominator of the relevant policies. The doctrinal elements, in different areas to a different degree, enter indeed as a part of the complex mixture that can be seen as the ideology of mercantilism and in which ideas of national power, wealth, and welfare are commingled partly as determining and partly as vindicating factors. But even under the most favorable circumstances the doctrinal elements are of subordinate significance, and some of their aspects are clearly in the nature of false pleading or "false consciousness" in the sense that they serve to disguise the actual aims of policies and do not present real programs or choices for action. This is most blatantly true of arguments for full employment, however eloquent mercantilistic writers may have waxed over the need to find "employment for our poor." For whatever view may be held of the analytical truth contained in those writings, they are not always remarkable for truth in the simple everyday meaning of the word. To repeat, doctrinal history, at least in this case, is not the proper avenue for approaching what after all was a long and eventful period in economic history, the timespan, as Heckscher put it, extending between the Middle Ages and the days of laissez faire.

I am sure I need not add that this conclusion regarding the theoretical writings of mercantilists—and their critique for that matter—has nothing to do with the examination of economic policies of the era, their course and their effects, with the help of modern tools of economic and quantitative methods. On the contrary, let me repeat that there is a crying need for such explorations. But they must be placed within relevant historical patterns, or in other words, must take their guidance from, and be conducted as, economic rather than intellectual history.

As one approaches the period of laissez faire, one has to walk much more gingerly. Naturally, I have no intention to read classical economics out of economic history. Its impact upon the course of economic events cannot and need not be denied. But it is important to understand both the limits of that impact and particularly the specific medium through which it worked.

Classical economics is, of course, a theory of economic development. In this respect it is quite unlike at least some of the economic theories that came into vogue in the last decades of the nineteenth century. Adam Smith's increasing division of labor and the widening extent of the market are dynamic factors of growth. So is his view on the formation of the labor force, so strongly emphasized by McCulloch who certainly anticipated List's ideas on productive forces by viewing dexterity, skill, and intelligence of labor to be acquired by education as a portion of national capital.[4] List's original contribution lay not in recognizing the dynamic nature of the transformation of the productive forces but in the role he assigned to the state in engineering that transformation. But it was accumulation of capital in the narrower sense of the term that was viewed as the powerful engine of economic growth. This was no doubt a great analytical achievement. Yet it was precisely here that a theory of economic development was conjoined with a theory of economic undevelopment. What Adam Smith saw as something that had not yet happened anywhere and might lie in a remote future, that is, a state of stagnation following a country's "acquiring a full complement of riches" in relation to "the nature of its soil, climate, and the situation with respect to other countries,"[5] became in the hands of his successors a much more imminent contingency. Capital meant growth, but "'industry [i.e., industrial growth] is limited by capital" in the phrase used by Bentham and John Stuart Mill,[6] and it came to be firmly believed that it would be so limited. The optimistic view of boundless progress, so important as a source of creative impulses and great expectations in development, was thus dimmed, if not obscured, by the vision of the ineluctable advent of stationary state. To be sure, the analytical meaning and value of the concept must be distinguished from its use as a prognostication of the future course of historical events.[7] But also that prognostication was arrived at by the use of analytical tools, such as rate of growth of population, falling profits, constant wages, and rising rents. And it is significant that it suffered from a fatal underestimation of technological progress. It makes, of course, no difference in this context whether the stationary state was viewed with dismay or with satisfaction. What matters is that the economic historian studying the period cannot view it as a brief chapter in the history of economic progress, as a desperate race against time. It is true that Ricardo at times spoke of a "progress in prosperity" that "would be beyond the power of imagination to conceive,"[8] if only the corn laws could be erased from the statute book. But the general power of the analytical model was so great that it is difficult to say that it was meant to apply to nothing else but a mature, insular economy. Nor was Malthus' theory of population seen in this narrow way. It would seem that it was rather the urgent desire to postpone, to stave off the inevitable that made the dry Ricardo to be so much out of character in his emotional appeal to the House of Commons: "I implore you to recollect that you are legislating for the happiness of millions and that there is no evil so intolerable as the high price of food."[9] If, by contrast, Cobden's and Bright's magnificent crusade was totally unburdened by the feeling that the fruits of the hard-earned victory would begin rotting away in their hands, it does mean that ideology succeeded in liberating itself from a theory to which in other respects it had owed so much.[10]

---

[4] J. R. McCulloch, *The Principles of Political Economy* (Edinburgh, 1843), pp. 117–19.

[5] Adam Smith, *The Wealth of Nations* (Modern Library, New York, 1937), pp. 94 and 96.

[6] John Stuart Mill, *Principles of Political Economy*, Collected Works (Toronto, 1965), Vol. III, p. 528.

[7] Joseph A. Schumpeter, *op. cit.*, p. 562.

[8] David Ricardo, *The Works and Correspondence* (Piero Sraffa, ed. Cambridge, 1925), Vol. V, p. 55.

[9] David Ricardo, *op. cit.*, pp. 183–84.

[10] How much precisely the repeal owed to economic theory is not easy to say. It was possible for Morley to write a two volume biography of Cobden without more than two brief and rather irrelevant references to Adam Smith. (John Morley, *The Life of Richard Cobden*, London, 1908, Vol. I, p. 119; Vol. II, p. 461.) And George M. Trevelyan in his long biography of John Bright refers to the problem in

The repeal of the corn laws in England stands out as one of the mighty monuments of the vast creative destruction to which classical economics had made its contribution. Nassau Senior once explained the reason for the prosperity of Britain to skeptical Frenchmen by saying: "It is the triumph of theory. We are governed by philosophers and political economists."[11] Clapham once said that there "had been a bit of Bentham in every great reform, destructive or creative"[12] although he did not say that the authors of the reforms were conscious of carrying out Bentham's precepts. And yet, one need not be surprised, as one looks through modern treatises on economic history, to find a general desire to play down the importance of the theoretical contribution. The following couple of quotations, selected at random, are fairly representative.

"The change in sentiment among the English industrialists is traceable not so much to a theoretical view of the advantages of a system of natural liberty as to the pressure of new forces and new conditions."[13] Or: "New capitalist classes began to demand greater freedom in the conduct of their enterprises. For this they found welcome support in the new economic doctrines of natural liberty and laissez faire."[14] At times the attribution of policies to economic theory was dubious to say the least. It was, for instance, Richard T. Ely who once referred to the abolition of serfdom in Prussia under Baron von Stein as a shining example of the influence of classical economics.[15] But the great reform of 1807 was so clearly the consequence of the smashing defeat Prussia had suffered at the hands of Napoleon and of the perilous domestic situation that resulted from it, that the well-established fact that Von Stein, after his years at Goettingen, had read Adam Smith rather pales into insignificance by comparison. I might add that the perusal of Adam Smith did not prevent Stein from advocating legislation that would protect the estates of the East-Elbian Junkers from the disruptive forces of the market.[16] And indeed, the problem is not a simple one.

Lionel Robbins, in his well-known series of lectures,[17] tried to defend the classical economics against what he called popular mythology which had accused them of narrow-minded opposition to any sort of social policies. This was in many respects a very convincing effort. But what matters is that along with a popular mythology that was critical of classical economics, there was also a large body of popular mythology that was favorable to it, this positive branch being probably just as far removed from the actual scholarly contents of classical doctrines as was its negative counterpart. It was then this popular mythology, that is to say, the ideology of laissez faire, that in a very real sense did become the spirit of the age, and something must be said both about its importance and its limitations.

It is not easy nowadays fully to ap-

---

one single paragraph in which he is willing to confine the influence of "sound theorists" to the early phase of the free trade movement, which ended with Huskisson. (*The Life of John Bright*, London, 1913, p. 55.)

[11] Nassau Senior, *Conversations with M. Thiers, M. Guizot, and Other Distinguished Persons During the Second Empire* (London, 1878), Vol. I, p. 169.

[12] J. H. Clapham, *An Economic History of Great Britain* (Cambridge, 1932), Vol. II, pp. 386–87.

[13] Bowden, Karpovich, Usher, *An Economic History of Europe Since 1750* (Boston, 1937), p. 86.

[14] Ernest Bogart, *Economic History of Europe* (London, 1942), p. 71.

[15] Richard T. Ely, "The Past and Present of Political Economy," *Johns Hopkins University Studies*, Second Series (Baltimore, 1884), Vol. III, p. 22.

[16] Wilhelm G. F. Roscher, *Geschichte der Nationaloekonomie in Deutschland* (Munich, 1874), pp. 716–17.

[17] Lionel Robbins, *The Theory of Economic Policy* (London, 1952).

preciate how pervasive the laissez faire ideology was and how deeply it affected minds whose interests were far removed from economics. Even parliaments were seen as a faithful replica of the free market. Free competition was bound to lead to preestablished harmony. Just so, truth, absolute truth, was to emerge from the free clash of opinions in the parliamentary debates. When the abortive Paris Commune issued an edict introducing rent control, it caused Gustave Flaubert to write furious letters about men who in their ignorance and stupidity believed that they could change the natural law of demand and supply.[18]

The laissez faire ideology rejected positive measures of governmental intervention in economic affairs by the formidable triad which declared them to be either superfluous, or harmful, or ineffectual. In so doing, it was strongly felt, it was informed by the last word of economic science, just as later the ideology of Marxian labor movements was felt to be the carrier of scientific socialism. I think it is fair to say that, as time went on, it was particularly the last element of the triad that became crucial; that is, the futility of attempts at government intervention to assert themselves against the work of economic forces. It implied the belief in the primacy of the economic factor, and also in this respect Marxism was a natural product of the liberal era (in the contemporaneous sense of the word). The analytical continuity between Marx and classical economics is paralleled by the continuity in broader social interpretations. It is only because Marxism alone has survived as a relic of a bygone age that nowadays ideas of economic determinism are so readily regarded as specifically Marxian. But if we try to express as briefly as possible the dominant idea of the period with regard to the relation between economics and politics, we might say that a curiously twofold view prevailed: on the one hand, what was done in the political sphere could not affect the economy, precisely because of its invincible resistance to interference; and, on the other hand, and most importantly, pursuit of correct economic policies was bound to lead to benign results in the sphere of politics, in particular banning the dangers of war and assuring international peace. That was certainly how Cobden viewed his commercial agreement with France—the great triumph of free trade. And rightly so, because by one fell blow the dread of imminent war, including the fear of invasion of the British Isles, was swept away, and the treaty of commerce became in fact a treaty of peace. The same ideas motivated the British government when as late as the summer of 1939 they still hoped that a broad commercial arrangement with Germany would assure the maintenance of peace—a futile attempt to solve the problems of the twentieth century with the devices inherited from its predecessor.

It is natural that the denial of the economic significance of political boundaries was an integral part of this complex of beliefs, the ideology thus going beyond Ricardian theory which assumed that factors of production did not cross state frontiers. Thiers was a protectionist, and a narrow-minded one at that, but when the moment for crucial decisions was at hand in 1871, he, too, proved a son of his age and lightheartedly surrendered French iron resources, because after all they could always be obtained through importations from Germany. And very similarly, Schumpeter followed this train of thought when in 1919 he appeared for the same reasons to view with equanimity the economic consequences of the disintegration of the Austrian Monarchy. For foreign trade—the economic force—again was expected to nullify political decisions. The view was not borne out by the facts, but it was not surprising in one whose formative years had passed in the nineteenth century. For the major shifts of European boundaries in the nineteenth century did not cause offense to the dominant ideology. German and Italian unifications clearly created more meaningful economic units, and the dismemberment of the Ottoman Empire, which, no doubt, did cause some economic losses, occurred too far away, on the fringes of Europe, to excite much notice and interest.

---

[18] Gustave Flaubert, *Correspondance* (Paris, 1924), Vol. III, p. 312.

But the spirit of the age is one thing; economic analysis is another. The laissez faire ideology was quite selective in what it did and what it did not accept from classical economics. The wage fund doctrine is an obvious example. It was greatly welcomed because it was said to show that increases in wages were "scientifically impossible." Distinctions between a short-run and long-run concept, or, even in the short-run, between the upper bound of availability and actual disbursements was not what ideology needed. Nor was it bothered by any connections between wage fund and agricultural output. As Schumpeter put it, "the public [read ideology] had caught hold of a surface meaning, and this was all it was interested in."[19] What mattered was that any attempt to raise wages could only benefit one segment of labor at the expense of another, and thus remain ineffectual as far as aggregate wages were concerned.

Consistency is seldom the virtue of ideologies.[20] It was not the theories of the classical school, but convenient fragments thereof—at times words rather than concepts—that entered the laissez faire ideology. And that was particularly true of the continent. Adam Smith's unpleasant remarks about the possible pernicious nature of high profits or about the detrimental impact of merchants and master manufacturers on public policies, including their inability to have a meal together without its leading to some conspiracy against public interest, were duly disregarded. McCulloch regarded prohibitions of workers' combinations as oppressive,[21] but popular mythology on the continent was certainly much more friendly toward laissez faire policies vis-à-vis business monopolies than vis-à-vis trade unions. It was only the latter that were usually stigmatized as being both ineffectual in the long run and harmful in the short run. While competition *of* labor was viewed with an auspicious eye, competiton *for* labor was another matter. The free labor market was said to insure the equality of the partners, provided, however, that a breach of contract by the master was to be a matter of civil law, while a breach of contract by the worker was to evoke sanctions by the penal code. Even more strikingly, tenets of economic individualism were easily combined with attitudes which tended to shape the relation in the factory between master and laborer in the image of the relation between the seignior and his serf.

On the continent, in a country like Germany, and not in Germany alone, laissez faire ideology absorbed without difficulty the ideas of economic nationalism. It is, therefore, not surprising that the actual flow of events was not necessarily in agreement with the spirit of the age. As an economic historian of today looks at the problem of economic significance of political boundaries, it appears that whatever the power of pure laissez faire ideology over the minds of men, an approach which uses the individual state as the primary unit of observation casts more light on the nature and cause of industrial change than alternative approaches focusing on large international trading areas. It is instructive to discover that, just as I was trying to show in the case of mercantilism, arranging the individual political entities in the nineteenth century according to the degree of their economic backwardness yields fairly consistent and intelligibly differentiated patterns of industrial development.

---

[19] Joseph A. Schumpeter, *op. cit.*, p. 670.

[20] Of ideologies, that is, viewed as a social phenomenon. No one will want to deny the admirable consistency of John Bright's beliefs and action. But a man of Bright's character is a *rara avis* in ideological skies.

[21] J. R. McCulloch, *op. cit.*, p. 261.

One of the reasons, albeit only one among several, was that the spirit of the age itself was liable to change. Whatever the insistence of the doctrine on the invincibility of economic forces, the doctrine itself did not prove invincible. Tariff protection and labor relations were outstanding areas. The dogma of ineffectuality of state intervention did not prevent, from a relatively early time on, introduction of measures of social legislation.

They were first introduced apologetically and described as exceptions to general rules of correct policy making. But gradually they came to be regarded, at least up to a point, as wholesome improvements upon the economic system. Some of the measures were clearly inconsistent with the tenets of classical economists in the sense of going beyond what they would have approved. Most of those measures were inconsistent with the ideology of laissez faire. But to say that they were introduced on the continent of Europe because of the appearance of a new school of economics—the German historical school which Richard T. Ely welcomed as an "occasion for rejoicing"[23]—would be first of all neglecting a large number of crucially important factors: First the fear and then the pressure of the rising labor movement; the anti-industrial sentiments of certain powerful groups within the social structure; competition for labor among industrial regions and branches; the aspirations of what in Germany was called *das soziale Koenigtum;* and the mechanics of forging parliamentary majorities. More important, however, in the present context is that the change was not due to the appearance of new economic theories. The German historical school, in its analytic helplessness, did not and could not provide any theoretical support for those measures.

As far as I can see, the only analytical contribution of the school was the use Brentano made of Hermann's critique in order to cast doubt—by somewhat irrelevant analysis—on the concept of the wage fund, without as Taussig says putting "anything very distinct in its place"[24] or rather, as we might say, replacing the wage fund doctrine by the practice of the British trade unions and the British factory legislation. The impressive empirical work performed in long decades since 1874 by the *Verein fuer Sozialpolitik* was so loosely connected with economics that one might say that what was being changed was not the theory, but the dominant ideology. This, too, was a case of economic policy *sans doctrine*. Something more will be said presently on the role of Marxism in this respect.

There was, however, an area with regard to which the belief in the ineffectuality of government intervention was held with particular persistence and rigidity. The reference is to the business cycles, those negligible intervals for which classical economics left no consistent body of theory, but for which the laissez faire ideology, aided no doubt by the growing literature on the trade cycle, had developed a policy it believed to be in the spirit of the masters. A depression with its falling wages and prices had to work itself out, producing its wholesome winnowing effects, enhancing the efficiency of the surviving enterprises, and reaching the stage where the resurgence of demand and the pressure of accumulated capital dispositions, combined with low interest rates, and possibly supported by fortuitous potentiality of technological innovations, created the natural turning point for the new upswing.

---

[23] Richard T. Ely, *loc. cit.*, p. 64.

[24] F. W. Taussig, *Wages and Capital* (London, 1896), p. 280.

I cannot possibly discuss here the general and complex question as to the role of new economic doctrines in bringing about a change in the relevant attitudes. But I should like to refer to the fateful economic policies pursued in Germany during the Great Depression of the 1930's, because they shed at least a negative light on the problem. Until a few years ago, I believed that those policies were essentially the result of the dominant traditional views. It is in these terms that they were contemporaneously presented and defended. The deflation had to run its course, and since rigid, because monopolistically held, prices and wages obstructed that natural course, they had to be forcibly cut by government decrees and at the same time budgetary expenditures drastically reduced. For the rest, there was nothing to do but waiting, not for Lefty or Godot, but for the Pigou effect—unconsciously, that is, and *ante litteram*. I also believed that this policy of deflation was powerfully reinforced by the collective memories of the German public whom the monetary disasters of the early 1920's had left with a neurotic dread of inflation. I have come to understand that neither received opinion nor fear of runaway inflation can be seen as the strategic factor determining the complex of policies which filled what probably has been the most dramatic chapter of modern economic history, its contents demolishing all our artificial boundaries between disciplines and becoming an integral part of general history—not of Germany, but the world. For there is little doubt that those policies opened the door of the Chancellery of the Reich to Adolf Hitler and must be regarded as the primary, and not so very remote, cause of the second World War. I must, however, warn that no positive inferences regarding economic determinism should be drawn from this statement, as I shall make clear shortly.

It is of course undeniable that orthodox economic views on depression policies were far from being extinct. They were held by members of the staff of the *Reichsbank*. Wilhelm Roepke, in advising the Central Bank, still spoke of the self-curing forces of the depression.[24] It is, furthermore, true and not paradoxical at all in the light of the preceding intellectual history that they were supported by orthodox Marxism. The social-democratic labor movement in Germany had moved a long way from its original position of alienated orthodoxy, the Marxian ideology having indeed become a "false consciousness" as had been clearly and candidly recognized and said by one of the outstanding leaders of the movement, Rudolf Hilferding.[25] But certain parts of the ideology were dormant rather than dead, and in a critical period they returned to life. Hence, despite occasional willingness to act as a physician at the "sickbed of capitalism," as the phrase ran, there was much talk about the depression inevitably being "the last capitalist crisis" and the imminent collapse of capitalism, as predicted by Marx or at least by Rosa Luxemburg.

---

[24] As related by the President of the *Reichsbank*, Roepke did so *"mit grosser Bestimmtheit"* (cf. the reference cited in footnote 30 below, p. 247). In private conversations, Professor Gottfried Haberler has suggested to me that the sentence in the text does not do full justice to Roepke's position and to the more positive views on depression policies which he expressed at that time. I think that this is quite correct and should be made clear, not only in deference to the memory of a fine scholar and a courageous and upright man, but also because a more complete presentation of ideas of an economist who was and remained strongly wedded to the creed of economic liberalism tends to support the main point of this section of the paper. In his short article "Ein Weg aus der Krise" (May, 1931) Roepke still spoke of the *Selbstheilungskraefte* of the depression, but at the same time he suggested the need of an *Initialsuendung* through credit expansion via a combination of long-term loans from abroad with a government investment program particularly in the area of road building and electric power. (The article is reprinted in Wilhelm Roepke, *Gegen die Brandung*, Zuerich and Stuttgart, 1959, pp. 54–60.) These ideas stemmed from the Second Report of the so-called "Brauns Commission." After the fateful collapse of the German banks in July, 1931, Roepke developed them with greater vigor in his much more extensive analysis of the Commission's reports ("Praktische Konjunkturpolitik, Die Arbeit der Brauns-Kommission," *Weltwirtschaftliches Archiv*, Vol. 34, 1931:II, pp. 423–64). He still reiterated his view that truly successful antidepression policy must be pursued during the period of the upswing and that nothing can and should be done during the "crisis proper"; that is to say, the period of economic cleansing and readjust-

And yet, it is most important to understand that along with the received opinion, there were also very different, much less passive and defeatist, views in circulation. Among them, no one was probably as important as Wilhelm Lautenbach, an economist of considerable analytical talents even though his posthumous glorification as a fully fledged Keynesian before Keynes's *General Theory,* no doubt, is an unnecessary exaggeration. Lautenbach understood clearly (as he wrote in 1931) that "the natural way to overcome an economic and financial state of emergency (*Notstand*) in the capitalist economy is not contraction but expansion."[26] And he also used to pronounce vigorously the elementary truth that inflation was not and could not be a danger in an economy afflicted by mass unemployment. In fact, even after the collapse of the Weimar Republic, he told the new chancellor of the Reich: "Herr Hitler, you are now the most powerful man in Germany; yet the one thing you cannot do, however hard you may try, is to bring about an inflation under the present conditions in Germany."[27] If even this statement contained some idea of ineffectuality of governmental policies, it was certainly of a very different kind, as is clear when it is compared with a thought Count Valuev, a Minister of Alexander II of Russia once (1861) expressed in a memorandum to the Tsar. He said: "Your Majesty can by one stroke of the pen abolish the whole code of laws of the Russian Empire, but you cannot raise by one kopek the exchange rate of the ruble at the Stock Exchange of St. Petersburg."[28] This was indeed the voice of laissez faire, ringing out clearly, the breadth of the continent away from the cradle of laissez faire doctrines.[29]

Lautenbach, a top official in the Economic Ministry—the *Reichswirtschaftministerium*—had to exercise moderation in his public utterances, but he was forceful, outspoken, and concrete in

---

ment. But he found bitter words about the business cycle policy of liberalism (i.e., laissez faire) whose consequences lead to *"wahrhaftigem konjunkturpolitischem Nihilismus"* (p. 450); and he emphasized that "a policy that attempted to counteract deflation had nothing at all to do with inflation" (p. 461). It is another matter that at least one-half of the recommendations made (the plea for large-scale, long-term capital imports) was hardly realistic in the conditions prevailing in 1931. But it is significant that even Roepke, despite his basic views, was moved to attack the laissez faire attitude of "nihilism" toward the depression.

[25] Rudolf Hilferding, "Probleme der Zeit," *Die Gesellschaft,* Vol. I: 1 (April, 1924), p. 5.

[26] Wilhelm Lautenbach, *Zins, Kredit und Produktion* (Tuebingen, 1952), p. 139.

[27] Wilhelm Lautenbach, *op. cit.,* p. X.

[28] Count P. A. Valuev, *Dnevnik, 1877–1884* (Diary) (Petrograd, 1919), p. 195.

[29] It is interesting to compare this nineteenth century statement with an assertion by Pososhkov who may be considered a Russian mercantilist "theorist" and who (in 1724) wrote that it was entirely within the power of the Tsar to impart to a kopek the value of a ruble. Pososhkov had no qualms about possible ineffectuality of governmental measures. Cf. I. T. Pososhkov, *Kniga o skudosti i bogatstve* ("A Book on Poverty and Wealth") (Moscow, 1951), p. 123.

his official memoranda, and he was excellently placed to let his views be known both to government and central bank. Under his guidance, ambitious plans for public works were being prepared, which, as Edgar Salin rightly notes, were subsequently generously drawn upon by Hitler's government in its full employment measures. Why then did the Bruening government fail to follow what surely was a most promising alternative to the sterile policies to which it remained so stubbornly attached? Now, after the publication of the memoirs by the late Hans Luther,[30] the president of the *Reichsbank* in those years, the reason has been authentically explained. It is simple and rather shocking. The government did not want an improvement in the economic conditions in Germany until the Allied powers could be induced to abolish the German Reparations. It wanted a deterioration. What mattered most was to reveal *ad oculos* Germany's inability to pay, and the deflationary policies were indeed excellently suited to fit that purpose. Once the goal was attained—as it was then, posthumously for the Bruening government, in the treaty of Lausanne—the plan was radically to cut the rudder and to embark upon policies of rapid reduction of unemployment. Thus neither defunct theories nor an ideology that was perhaps not quite defunct, although it should have been, but an eminently political consideration, altogether informed by nationalistic aims, appears to have guided and misguided the policies. I feel that in writing —or rather rewriting—this momentous chapter, an economic historian has little reason to place much emphasis on the role of past economic theories, whether in pure form or diluted as they may have been within a broad ideological substance.

Returning to the nineteenth century and going beyond, and away from, classicial economics, it should be clear that in dealing with the industrial history of the nineteenth century in its relation to doctrinal history neither the name of Marx nor that of St. Simon can be avoided. The influence of St. Simon upon industrial history, to my mind, considerably exceeds that of Marx. True, Schumpeter once asked: "Can the reader be so thoroughly imbued with intellectual misconceptions as to believe that, without St. Simon's teaching, the Crédit Mobilier would not have been founded and managed exactly as it actually was, and would not have crashed exactly as it actually did?"[31] I am not at all opposed in principle to raising reasonable counterfactual questions. And the main tenor of this lecture should safeguard me against the intellectual conception or misconception of attributing too much weight to intellectual history. But I do know that the St. Simonian impact extended far beyond the relatively short-lived history of the Crédit Mobilier. In saying that I am not thinking so much of the intellectual impact of the St. Simonians upon the optimistic view of economic development as formed by Friedrich List, in whose hands St. Simonian socialism was replaced by a nationalist ideology. I am thinking of economic institutions, and I do not see how the history of the continental investment bank—one of the great innovations of the century—could be written without paying due attention to the St. Simonian spirit which informed and motivated some of the pioneers of the new banks. To be sure, there had been Jacques Laffitte and there had been the Belgians, both antedating the St. Simonian ventures, but neither the one nor the other was able in the least to exert any comparable influence. If Walter

---

[30] Hans Luther, *Vor dem Abgrund, 1930–33, Reichsbankpraesident in Krisenzeiten* (Einfuehrung von Edgar Salin, Berlin, s.a.).

[31] Joseph A. Schumpeter, *op. cit.*, p. 462.

Bagehot once wrote that it was impossible to change the banking system of a country by writing books about it, it would seem that St. Simon and St. Simonism came pretty close to being a refutation of that proposition. At the same time, however, Schumpeter is quite right in mentioning St. Simon's "prejudice against economics"[32] and in saying that St. Simon's contribution to economic analysis was nil, even though he adds that keener theoretical minds might have profitably tried to build St. Simon's concept of entrepreneur into analytical schemes. By contrast, Schumpeter was willing to assign a place for St. Simon in the history of economic ideas, but that concept as used by Schumpeter was so infinitely broader than any reasonable concept of history of economic doctrines as to come rather close to what I have meant when speaking of ideology. At the same time, it must also be recognized that the fully developed investment bank of the German type was the result of a long learning process, from which the two fundamental features of the institution emerged: (1) the comprehensive scope of the bank as a department store in banking and (2) the use of the short-term open account for long-term industrial investment, coupled with the practice of mutual aid among banks in case of emergencies. As such the investment banks were so far removed from, and superior to, St. Simon's original ideas and the early St. Simonian practice that it is necessary to remain aware both of the force of the original impact and of its historical limits. It must be added, however, that the new banks were altogether unsupported either by past or by contemporaneous economic theories. In fact, their policies were under continual attack by economists who never tired of predicting imminent and ineluctable collapse of the banking system that was pursuing dangerously unsound policies. It is incidentally only another sidelight on the inconsistencies of ideology that the individualistic entrepreneur on the continent who stressed the need of being absolute master in his house with regard to labor willingly accepted the role of the bank in codetermining both entrepreneurial and managerial decisions.

When it comes to asking the same question with regard to Marx, the problem of his impact on economic history presents itself in a somewhat different guise. However one views Marx as an economist, be it as creator of a powerful dynamic system or as a minor Ricardian, no one denies that he occupies a firm position in the history of economic doctrines under any interpretation of the concept. There is furthermore no doubt that Marxian theories became accepted by mass movements which again indubitably exerted their influence on the course of economic events. Marxian theory of value became writ large over the banners of huge political parties. It was well internalized by millions of adherents of those parties and, of course, was raised to the level of absolute truth in Soviet Russia. And yet, I think it is correct to say that the total impact of Marx as an economist was very small, if not negligible. What was accepted was the implied ethical message of the doctrine rather than the doctrine itself. For the practice of Marxian labor movements the fatal inconsistency between Volume I and Volume III of *Das Kapital* was exquisitely irrelevant. No one cared whether the theory did or did not explain the structure of relative prices. Equally irrelevant was Marx's metaphysical derivation of labor value, so strongly reminiscent of Descartes' method of discovering—or rather asserting—extension as the essence of objects. What mattered was that the concept of surplus value with its rate of exploitation appeared intuitively in accord

[32] *Ibid.*

with, and well justified by, the discontent of men whose standard of living was distressingly low and who felt deeply about the inequality of distribution of income and wealth, as people generally do as long as levels of productivity are relatively low and the national product is relatively small but growing. The theory of surplus value was supported, not by economic analysis, but by the conditions of the age. And if from the theory of creation and distribution of surplus value emerged not only the concept of economic classes but also the assurance that the unjust economic conditions would be eliminated largely by the very operation of the economic system which then by one easy push would be replaced by a new system, an incarnation of goodness and justice—this, too, fitted to a nicety the spiritual needs of those who felt exploited, discriminated against, and helpless. It was indeed comforting, in the specific circumstances of the nineteenth century, to receive such an assurance from what was considered the last word of economic science. Schumpeter, as far as I am aware, never argued that the history of the European labor movements would have been exactly the same if Marx had never written, and it is difficult not to assign to Marxism a place of high importance as a factor contributory to the formation of those movements. But, to repeat, what was received and worked as intellectual yeast was not the theory in any reasonable sense of the word, but the ethical norm in conjunction with a gospel-like promise of delivery. All the talk about scientific socialism could not change the fact that what was created was an ideology which contained, along with other things, para-theoretical elements and did so probably to a somewhat higher degree than most ideologies.

The labor movement so formed and imbued with these ideas came to play a not inconsiderable part in economic history. First of all, negatively, because, whatever Marx's teaching on technological unemployment, the errors of the compensation effect, and the industrial reserve army, what ideology had distilled and accepted as an axiom was that technological progress was a good thing in terms of "1066 and all that," because every step of capitalist progress was bound to hasten the advent of socialism. As a result, Marxian labor movements did not impede industrial growth by "mechanoclastic" episodes. And, second, because of the subsequent indubitable impact of the labor movement upon passage of acts of social legislation and its pressure, through labor unions, for rising wages and improvements in working conditions. But the point is that in so doing they were, in fact, refuting the theory and abandoning the ideology to which by inertia, tradition, and possibly correct appraisal of psychology and sociology of mass movements they continued to adhere.

But what has been said so far refers largely to labor movements in Central Europe. In Russia of the 1890's it was left to Marxism to perform, in addition to its organizational function in forming the social-democratic party, a specific role as the ideology of the contemporaneous industrialization. There is little doubt, I believe, that in this area the contribution of Marxism was important indeed. It helped to overcome the inveterate populist attitudes of the intelligentsia, and beyond that it even affected certain influential segments of the imperial bureaucracy. To have made industrialization acceptable to the intelligentsia—and that meant to crucial segments of public opinion—was no mean achievement. Again, to an extent which I cannot measure and which defies counterfactual needling, Marxism affected the course of economic history. But here again I find it difficult to say that it was Marx as an economic theorist that produced the effect. To be sure, Marx's economic theories were read eagerly, mostly in Karl Kautsky's well-known popularization, but often enough in their trebly voluminous form; and to have read them in the original was a goal aspired to by many and attained by some. And yet, the theory was hardly more than the background for a single proposition: to wit, that capitalism was inevitable everywhere, and hence also in Russia, and that capitalism was the certain road leading to socialism. Did this proposition actually emerge from *Das Kapital*? It did and it did not. For the populists whom the Russian Marxists confounded so effectively also knew and cherished their Marx; except that, doing possibly equal, if not greater, justice to the thought of the master, they managed to extract from it the opposite proposition; viz., that a latecomer like Russia, being unable to compete in foreign markets, would necessarily see its attempts toward capitalism founder on the shallows of insufficient aggregate demand. In fact, as far as the influence from theory went, you could take your pick, which is just another way of saying that it was ideology rather than theory that was at work in Russia of the time and exerting its effects upon economic history. It is still another matter that there was something unrealistic about all these logomachies. While the debates proceeded, the question whether Russia should and would follow the road of industrialization and whether such industrialization could continue unabated had been long solved convincingly by the very course of economic events. As so often in history, intellectual history was straggling behind the times, executing what in Russian military drill is called a "step on the spot" (*shag na meste*) which, I feel, is rather more graphic than the English equivalent of "marking time." At any rate, such connection as can be ascertained between Marxism and Russian industrialization remained altogether on the surface. No one could claim that Marxian theories influenced in any way the manifold concrete measures of governmental policies.

I have expressed elsewhere my views on the role of Marxism in the economic history of the Soviet Union, dealing not only with Marxian economics but also with the other ingredients of Marxism. Only the former should be mentioned here, doing it as briefly as possible. Since Marx deliberately abstained from issuing any blueprints for the socialist economy of the future, the leaders of the Soviet government, who believed in the essential difference between capitalist and socialist economies, had nothing but the few generalities of Marx's *Critique of the Gotha Program* to go by. From that they could deduce the undesirability of an egalitarian wage structure at the early stages of socialist economy, but the actual structure of wages, i.e., the degree of inequality, was something for which Marxian theory equipped them only with the nonoperational, if not question begging, concept of socially necessary labor. At any rate, they were quite willing at one point to introduce general equality of wages and were diverted from that course not by a rereading of Marx's *Critique,* but under the pressure of formidable practical necessities. Interest rates were indeed regarded as "un-Marxian," implying recognition of the productivity of capital. But they had no inhibition in charging interest on short-term loans, and although until very recently, long-term capital was made available in the form of interest-free grants, cost calculations of investment projects and comparison among them did operate with interest rates, even though in a crude and inchoate fashion. From the tenor of Marx's pronouncements on centralization and concentration came unwillingness to recognize the phenomenon of diminishing returns. For a while, the sky seemed to be the limit as far as the optimal size of plant and enterprise was concerned. But when for economic and political reasons the unwisdom of this attitude became clear, they had no hesitation to decry it as gigantomania and to attribute it to the cabals of the enemies of the people. I believe that the economic order (or disorder) as was developed in Soviet Russia was created, not in obedience to any theoretical tenets, but as a pragmatic response to the exigencies of the practice with the power mechanics of the dictatorship well in mind. Thus my conclusion is that hardly anything in the momentous story of Soviet economic policies needs, or suffers, explanations in terms of its derivation from Karl Marx's economic theories. The Soviet continual reference to Marxism must be seen as primarily designed to vindicate policies which had little to do with their alleged fount and origin and clearly belonged in the sphere of ideology rather than theory. It is indeed with the help of Marx, although not Marx alone, that the nature of these techniques of persuasion and domination could be explained. But the desire to apply Marxism to the study of real or reputed Marxism was never visible within Soviet Russia—and rather rarely without it.

In the foregoing I have dealt on the whole with large, possibly overly large, problems. I cannot, within the scope of this lecture, address myself to more specialized areas of European economies in the nineteenth century. If I had, a somewhat variegated picture would have resulted, depending on the individual country and the segment of the economy. The reform of the Poor Law in England in its relation to economic analysis was surely different from the German regulation of the Stock Exchange in the same respect. Nevertheless, I should be willing to record, in a guarded way, my general impression that my conclusions would have varied to some extent, but not fundamentally, had I considered policies of central banks, or goverment policies with regard to central banks, or introduction of gold standard, or location of industries, or, naturally, tariff protection. With regard to the latter, such theoretical propositions as could be advanced in favor of protection were, I believe, irrelevant. This was obviously the case in Germany, where the infant industry argument was used to protect the senescent, if not senile, East Elbian grain growing and the giants of the iron and steel industry, neither of which could possibly qualify as infants, and the honor of being the spiritual father of this type of protectionism was foisted upon a rather innocent List. In this case, the whole body of relevant writings, including the records of parliamentary debates, does indeed make it possible to hazard a counterfactual proposition. If List, or his predecessors, had never lived, the course of German protectionism would not have changed by one jot. And it is equally clear that German economists, such as Wagner, Oldenberg, Pohle, and others, who raised such a loud cry in favor of protection, were writing not *qua* economists, but as political journalists. Nor am I ready to assume that it was the logical power of theoretical arguments in favor of free trade rather than the force of countervailing interests that prevented protection from attaining even higher levels.

As one looks back upon the economic history of Europe in the nineteenth century, one is tempted to paraphrase Axel Oxenstjerna's adage and to say: *Quantula theoria economica mundus regebatur*. But I am far from arguing that the fault is all on the side of the practice. Had the theories been more operational, primarily in the narrow, pragmatic sense of the word, had there been, that is, a sufficiently large stock of empirical knowledge to support them, their impact upon events would have been much greater.

We live now in a new era. Today's profession, whatever its more esoteric leanings, works with new and much more operational tools, and it has at its service the vastly increased body of primary quantitative information—the fruit of a formidable empirical effort. This body of information has been organized with a conceptual clarity that was altogether unknown to the earlier periods, and it is capable of being treated with the modern, and in fact very recent, methods of statistical analysis. On the other hand, both the shattering experience of the Great Depression of the 1930's and, particularly, the postwar interest in economic development have created, on the intellectual level, a rapprochement between economists and economic historians, and, on the level of practical policies, a greatly increased willingness to listen to economists' opinion and advice.

The mostly negative picture which I have offered in this paper is surely one-sided, and I am certain that I myself would be quite willing to listen with sympathy and approval to a more positive treatment of the same subject. Still, ill-balanced as my presentation has been, I think there may have been some merit in going over the bits of evidence that can be mustered in the service of a negative approach. But if the past may look bleak, the future is bright. The economic historians of future times in dealing with our, in many other respects so unhappy, century will be forced to stress the strong impact of economic doctrines upon the course of economic events. And perhaps this is precisely what will be done one night between Christmas and New Year, in the year of the Lord 2068, by another economic historian in another lecture given in honor of Richard T. Ely.

THEORY OF INNOVATION

AN ECONOMIC THEORY OF TECHNOLOGICAL CHANGE*

By William D. Nordhaus
Yale University

I. *Introduction*

It is commonplace to assert that most of the growth of output is due to technological change. It is more accurate to state that the growth of inputs cannot explain most of the growth of output. At the present time there is no compelling empirical evidence pointing toward technological change rather than associating increases in productivity with economies of scale, learning by doing, errors of measurement, or even sunspots. The failure to explain the residual may be due to the absence of an adequate theory—especially microtheory—of the generation and transmission of new knowledge. Once a reasonable model of the inventive process has been developed and tested, it should be easier to sort out the sources of productivity change among the different claimants. The present paper is a crude attempt to accomplish this task. The first five sections of the paper present an economic model of the inventive process and discuss its properties. Section VI then applies the model to two industries in order to test its predictions.

II. *The Setting of the Problem*

Although technology has long been recognized as an important part of the economic scene, invention has not been fully integrated into economic analysis. (Invention will be used as a general term for activities which expand the level of technical knowledge.) In fact, the "Schumpeter tradition"—treating invention as an exogenous force acting on the economic system—has led to widespread neglect to fundamental relations between invention and economic activity. Following the lead of the late Jacob Schmookler, important connections have been uncovered, indicating that over half patented inventions are assigned to profit-oriented corporations and that time series on patents show striking cum-cyclic correlations. This evidence has led most analysts to abandon the Schumpeter tradition.

Recent work integrating invention into conventional economic analysis (such as [1] [4] [11]) has given us a deeper understanding of the important problems in the economics of invention. These studies have highlighted: (1) the high degree of uncertainty residing in the outcome of inventive activity; (2) the public good character—or inappropriability—of inventions, except under extreme legal arrangements like a patent system; and (3) the indivisibility of invention, meaning that once a new process has been discovered it can be spread to all firms at (virtually) zero marginal cost.

It is essential that these distinctive properties of information be included in a serious model of the inventive process. In general, models investigating productivity (such as Lucas [5] and Griliches and Jorgenson [2]) treat inventions like any old capital good, with no problems of indivisibility or inappropriability. If the work mentioned in the last paragraph is taken seriously, then the capital-theoretic

---

* The author would like to thank Joseph Stiglitz for helpful comments. Work was done with the help of grants from the National Science Foundation and from the Ford Foundation.

models of invention do not capture the important characteristics of invention.[1]

The model analyzed and tested here attempts to cope with the problems of invention in a way quite different from the capital-theoretic view. The scenario is roughly as follows. The economy is neoclassical in all regards except in the production of knowledge. There are two conventional factors of production, capital and labor, which are combined to produce output by an aggregate production function. Labor, capital, and output markets all behave competitively.

There are several new twists which are introduced to generate the model of invention:

1. Inventions are produced within the system. An invention is viewed as a new process of production, or as a new vector or input-output coefficients.

2. Any invention is potentially a public good in the sense that it is indivisible, or that it can in theory be used universally at zero marginal cost. In practice the spread of inventions can be impeded for some time by inventor secrecy or by legal arrangements such as exclusive rights to trade secrets or to patented inventions. In the model we assume that the inventor has exclusive rights to use and/or license the invention for T years, after which the invention enters the public domain as public knowledge. T can be interpreted as the life of a patent, the average lead-time, or the length for which secrets can be held.

3. The most important point is that from an economic point of view the inventor has a monopoly over the invention for T years. He can either produce output or license the invention to other producers. After T years, his monopoly vanishes, and anyone can use the invention.

It should be stressed that the monopoly over information is essential for a sensible treatment of invention when invention is a public good. There is no more sense to having a "competitive" market for a blueprint for the xerography process than a competitive market for national defense.

In sum, the model used here is a traditional neoclassical model except for the introduction of a multitude of temporary little monopolies on information. It resembles in many ways the Schumpeterian view of the process of capitalist growth, with inventors taking the role of entrepreneurs and getting all the surplus. Instead of entrepreneurs getting the profits from introducing innovations or "new combinations" or innovations (as in the Schumpeterian scheme) inventors get profits from inventions. As will become apparent, this changes the shape of the capitalist economy in some interesting ways.

III. *The Model*

In setting down the model we will give first the conventional framework to be retained—equations (1) and (2)—and then give the technology and market for inventions—in equations (3) through (10). Important variables are as follows:

A = Level of productivity
C = Consumption
D = Cost per invention
I = Net investment
K = Capital stock
L = Labor
N = Number of inventions
p = Price level
q = Price of capital service
Q = Output
r = Interest rate
s_i = Per unit royalty on ith invention
S = Total royalties per unit output
R = Total inventive inputs
T = Life of monopoly on invention
V = Value of invention
w = Wage rate

---

[1] One of the illustrative shortcomings of the capital-theoretic model is that the competitive solution **is** Pareto optimal **as** long as institutions are taken **as** given. This is seriously at variance with the conventional wisdom.

s = Aggregate savings rate
α = Output elasticity of capital
β = Elasticity of $\hat{A}$ with respect to N
γ = Elasticity of $\hat{A}$ with respect to A
μ = Share of inventors
π = Rate of population growth
ϕ = Discount factor
$\sigma = 1/(1-\beta)$

Output is produced by a Cobb-Douglas production function:

$$Q = AK^{\alpha}L^{1-\alpha} \tag{1}$$

where A is the level of Hicks-neutral technological change.[2] Output is devoted to three uses: consumption (C), net physical investment (I), and research or inventive inputs (R). We assume that a constant fraction of income, $(1-s)$, is consumed. Labor grows exponentially at rate π. Since all conventional markets are competitive, factor rewards are equal to marginal physical product times net price; i.e., market price less royalty payments.

Since the net rental on capital is $q=rp$, the equilibrium level of capital is given by

$$K = \frac{\alpha Q}{r}\left(\frac{p-S}{p}\right). \tag{2}$$

So much for the conventional markets. We now examine the market for invention. First, we find the level of royalties for a given profile of past inventions. This is quite simple under the assumption that inventions are technically independent. More precisely, we assume that the absolute contribution per invention to total productivity, A, is independent of whether other inventions are used. This implies that the ith (of m inventions) contributes ΔA_i to productivity independently of whether inventions $j \neq i$ are used.[3]

Under this condition it is easily shown that the royalty on the ith invention as a proportion of market price equals the ratio of the change in productivity due to the ith invention to level of productivity:[4]

$$\frac{s_i}{p} = \frac{\Delta A_i}{A} \tag{3}$$

Using (3) we can derive an extremely interesting result. Add together the per-unit royalties for all unexpired inventions; i.e., all inventions less than T years old (call them inventions $1, \cdots, m$) and let S be the sum of royalties on these inventions. Using (3) we have:

$$\frac{S}{p} = \sum_{i=1}^{m} \frac{s_i}{p} = \frac{\sum_{i=1}^{m} \Delta A_i}{A(t)} = \frac{A(t) - A(t-T)}{A(t)}.$$

The share of inventors, μ, is given by:

$$\mu = \frac{SQ}{pQ} = \frac{S}{p} = \frac{A(t) - A(t-T)}{A(t)} = 1 - \frac{A(t-T)}{(At)}. \tag{4}$$

---

[2] Some cheap generalizations can be made as follows. If production is not Cobb-Douglas, the asymptotic results hold if technological change is Harrod-neutral and the elasticity of substitution is less than one. The instantaneous equilibrium in this case becomes much more complex. If the elasticity of substitution is greater than one the system explodes for any technological change.

[3] This assumption of no joint production is particularly bothersome. If inventions were not independent but inventors acted independently, the problem would disappear. The problem is that with interdependence a game theory problem of royalties arises. When inventions are independent the Nash equilibrium—the solution used here—is composed of dominant strategies and is thus unobjectionable. When inventions are not independent the Nash equilibrium is no longer dominant.

[4] This proposition was first shown by Arrow [1], and has been expanded in [9]. A demonstration runs roughly as follows: $\Delta A_i/A$ is the percentage cost reduction which can be attained by a firm using the ith invention. The total cost reduction including royalty is thus

$$\left(-\frac{\Delta A_i}{A} + \frac{s_i}{p}\right)p.$$

Putting (3) into this we see that s_i is the maximum

The distributional theorem in equation (4) is remarkably general. It holds for any production function and any past profile of inventions. It tells us that the share of inventors is the ratio of the increase in technology over the last T years to the actual level of technology. If a is the annual growth rate of technology, then the share of inventors is $1-e^{a-T}$, or somewhat less than aT.[5] Equation (4) looks almost like the statement that the compensation of inventors is equal to the marginal product of invention, but that is not correct. It is rather that inventors capture all the surplus for the life of the invention.[6] In fact, noninventors get at time t exactly what could have been produced with the technology of period $(t-T)$; inventors get all the extra output due to the increase in technology over the past T years.

Two further assumptions are necessary. First, the supply of inventors is taken to be perfectly elastic at the going wage rate. This is relatively harmless, and an extension to scarce, competitive inventors would be simple. The crucial assumption for all this analysis—and an assumption which cannot be swept under the rug—is that there is a relation between the amount of inventive input and the rate of technological change. Put differently, there is a reasonably well-defined production function for technology. This is a questionable assumption. On the one hand, it is difficult to point to any laws of production of technology. On the other hand, the fact that profit-oriented firms are becoming more involved in research would indicate that they feel there is a positive relation between inventive inputs and technology. Moreover, there is some scattered empirical evidence of such a relation between research inputs and productivity (see [7]).

Formalizing these assumptions, we can write the rate of technological change as a function of the number of inventions. We use the simple log-linear function:[7]

$$A/A = N^{\beta}A^{-\gamma} \tag{5}$$

Equation (5) states first that the rate of technological advance is an increasing function of the number of inventions produced at a given time, N. We assume, following arguments of Machlup [6] and others, that for a given initial stock of knowledge, A, inventions add less and less to the level of technology—thus diminishing returns to N. The other feature of equation (5) is that the rate of advance may be a function of the level of knowledge. Thus, if knowledge has advanced very far, then (for a given N) the rate of advance might be lower. The opposite might also be the case. In the dynamic analysis it can be shown that for stability A must be a retarding force on technological advance, thus $\gamma>0$.

We can now calculate the profit of inventors. The Nth invention gives an increase in productivity, $A'(N)$, given by

$$A'(N) = \beta N^{\beta-1}A^{1-\gamma} \tag{6}$$

Since from (3) the royalty, $s_N(t)$ equals the rate of productivity change and setting D equal to the cost per period per invention, we have the profit of the marginal invention, $V(N)$, equal to

---

royalty that an inventor can charge. It can be shown rigorously that as long as A_t/A is small (3) will also be the profit maximizing royalty for the inventor.

[5] This is trivial: Let $K^{\alpha}L^{1-\alpha}=F$. Then $(1-\mu)A(t)F=[A(t-T)/A(t)]A(t)F=A(t-T)F$.

[6] Equation (4) can be applied to show (*a*) that there is generally a bias in estimating output elasticities from relative shares. Thus if all inventor's royalties accrue to "capital," then the observed share overestimates capital's share by $(1-\alpha)[1-\exp(-\hat{A}T)]$; (*b*) in this case using observed weights underestimates $\hat{A}$ by the factor $\exp(-\hat{A}T)$; (*c*) as invention moves from the personal to the corporate sphere, capital's observed share will rise significantly.

[7] We have shifted to continuous analysis and will do so for the next two sections. This is inessential for the argument. Although continuous analysis causes some problems of understanding, the interpretation of discrete numbers of inventions and discrete changes in A gives the usual view.

$$(7)\qquad V(N) = \int_0^T s_N(t)Q(t)e^{-rt}dt - D.$$

The first term is the per unit royalty for the Nth invention times the number of units discounted back by a constant interest rate for the life of the monopoly. The second term is the cost. Putting (3) and (6) into (7), and normalizing by setting $p=1$ (thus converting r into a real interest rate), we have:

$$V(N) = \int_0^T \beta N^{\beta-1}A^{-\gamma}Qe^{-rt}dt - D.$$

If inventors assume that output and technology will grow in the future at their present rates of growth, $\hat{Q}$ and $\hat{A}$, the equilibrium becomes:

$$(8)\qquad V(N) = \beta N^{\beta-1}A^{-\gamma}Q\phi - D$$

where ϕ is a capitalization factor given by:

$$(9)\qquad \phi = \frac{1 - \exp[(\hat{Q} - \gamma\hat{A} - r)_{Z}T]}{r + \gamma\hat{A} - \hat{Q}}$$

Inventor equilibrium is attained when $V(N)=0$; i.e., when the marginal inventor makes zero profits.[8] Therefore equilibrium in the invention market implies:

$$(10)\qquad N = \left[\frac{\beta QA^{-\gamma}\phi}{D}\right]^{1/(1-\beta)}$$

This structural equation for the demand for inventions bears remarkable similarity to Schmookler's intuitive description of the determinants of inventive activity. A larger industry will induce a larger flow of inventions—this is a consequence of the indivisibilities associated with knowledge. Schmookler concurs, writing in a chapter entitled, "The Amount of Invention is Governed by the Extent of the Market . . . ":[9] "The most important relation for our purpose, however, is the fact that given the expected market share and the expected cost of invention for each possible invention, the number of machines it will pay to invent will vary directly with S, the expected size of the market." Schmookler did not investigate the other variables, although cost (D) and prior knowledge (A) did enter into his discussion.

IV. *Instantaneous Properties of the System*

We now have enough information to describe completely the economy. We will first discuss the instantaneous equilibrium. We can reduce the system to five variables, Q, A, N, π, and K, defined by the behavior equations (1), (2), (5), and (10), and the savings assumption:

$$(11.a)\qquad Q = AK^{\alpha}L_0e^{\pi(1-\alpha)t}$$

$$(11.b)\qquad ND + I = zQ$$

$$(11.c)\qquad K = \frac{\alpha Q}{r}(1 - \mu)$$

$$(11.d)\qquad A/A = N^{\beta}A^{-\gamma}$$

$$(11.e)\qquad N = \left(\frac{\beta QA^{-\gamma}\phi}{D}\right)^{1/(1-\beta)}$$

Differentiating all but the (11.b) logarithmically with respect to time, and letting hats over variables represent proportional rates of change (thus $\hat{x}=\dot{x}/x$):

$$(12.a)\qquad \hat{Q} = \hat{A} + \alpha\hat{K} + \hat{\pi}(1 - \alpha)$$

$$(12.b)\qquad NC + I = zQ$$

$$(12.c)\qquad \hat{r} = \hat{Q} - \frac{\hat{\mu}\mu}{1 - \mu} - \hat{K}$$

$$(12.d)\qquad \hat{A} = \beta\hat{N} - \gamma A$$

$$(12.e)\qquad \hat{N} = \sigma\hat{Q} - \sigma\gamma A + \sigma\hat{\phi},$$

where $\sigma = 1/(1 - \beta)$.

The important features can be seen by examining the (12.a) and the combination of (12.d) and (12.e):

---

[8] This equilibrium assumption implies some kind of priority in discovering new techniques. An alternative assumption might argue for all inventors averaging zero profits. In this case the equilibrium condition is (10) with the β inside the bracket replaced by 1.

[9] [11], p. 115.

$$(13) \quad \hat{Q} = \widehat{A} + \alpha\hat{K} + \pi(1-\alpha)$$

$$(14) \quad \hat{\hat{A}} = \beta\sigma\hat{Q} - \gamma(1+\beta\sigma)A + \beta\sigma\hat{\phi}$$

These equations clarify the interdependence between the growth of output and the growth of knowledge. Knowledge growth affects output by the conventional additive term, $\hat{A}$, in equation (13). The new twist is that this feeds back through the knowledge market, increasing the acceleration of knowledge by the factor $\beta\sigma$.

A second important interaction can be seen by examining equation (11.c). Recalling from (4) that $(1-\mu)=A(t-T)/A(t)$—which equals $\exp[-\hat{A}T]$ in equilibrium—we see that a rise in the rate of technological change depresses the capital stock by cutting into the post-royalty marginal product of capital. This is a result which does not seem to have been noted before. Rewriting (11.c) we have the equilibrium capital-output ratio:

$$K/Q = \alpha \exp[-AT]/\mathrm{r}$$

If, as empirical evidence indicates, interest rates remain roughly constant and technological change speeds up, then the equilibrium capital-output ratio should be declining.

V. *Asymptotic Properties of the System*

The economic system described in equation set (12) is rather different from the standard neoclassical model. The chief differences in structure are that technology can be interpreted in the growth equation as a capital good with unitary exponent, and that the technology-formation function is quite strange compared to the capital-formation equation. This strangeness is due to the fact that technology is subject to monopoly pricing rather than competitive pricing.

A long-run balanced growth path can be found as follows. Recalling from (11) that $\mu\hat{\mu}/(1-\mu)=\hat{A}(t-T)-\hat{A}(t)$, substituting this and (12.c) into (12.a), we have (15). (16) is from (14):

$$(15) \quad \hat{Q} = A(t) + \pi + \frac{\alpha\hat{A}(t-T)}{1-\alpha} - \frac{\alpha\hat{r}}{(1-\alpha)}$$

$$(16) \quad A = \beta\sigma\hat{Q} - \gamma(1+\beta\sigma)A + \beta\sigma\hat{\phi}$$

Consider paths with constant interest rates, $\hat{r}=0$. It is easily verified for such paths that the long-run equilibrium is reached when $\hat{\hat{A}}=\hat{\phi}=0$, so from (16) and recalling $\beta\sigma=\beta/(1-\beta)$:

$$(17) \qquad \hat{Q} = \frac{\gamma}{\beta} A$$

Putting this into (15) for $\hat{\hat{A}}=0$ implies

$$(18) \qquad A = \frac{\pi\beta(1-\alpha)}{\gamma(1-\alpha)-\beta}$$

It can be shown that a necessary condition for (17) and (18) to be a stable equilibrium is that the denominator is positive, so $\hat{A}$ will always be positive.[10]

Equation (18) has a few remarkable features. First, it is linear in population growth. If population is stagnant, then eventually technology will stagnate. The reason for this peculiar, nonneoclassical feature is that since $\gamma<0$ the annual number of inventions must be growing to keep up with the diminishing returns in the knowledge function. But the number of inventions can grow only if income is growing faster than $\hat{A}$. Thus unless population growth gives a stimulus to growth of income, both income and technology will eventually stagnate.

This peculiar result is one of several due to the more general feature of the model, that it displays aggregate economies of scale. As is obvious from the production function, doubling all "factors"—A, K, and L—quadruples the level of output. One of the nice features of the present model is that it allows this realistic feature without introducing inconsistencies into the market structure. Such large increasing returns are consistent with profit maximization because information is temporarily monopolized.

---

[10] Stability analysis of the differential-difference system in (12) is quite difficult. So far we have only succeeded in showing stability for the system which has a constant interest rate at the equilibrium value.

Other magnitudes of the system are easily found. From (12) and (17) it is easily seen that $\hat{N}=\hat{Q}$. Therefore the shares of output devoted to investment and invention are constant. We can find the interest rate by combining (11.d) and (11.e), obtaining

$$\frac{DN}{Q} = \beta\phi\hat{A} \tag{19}$$

So the interest rate is given by the solution of (19), (11.b), and (11.c):

$$z = \frac{DN}{Q} + \frac{I}{Q} = \beta\phi\hat{A} + \frac{\alpha\hat{Q}e^{-\hat{A}T}}{r} \tag{20}$$

This does not seem to be tractable. Since $(dz/dr)<0$, it is easily seen that there exists a single value of r satisfying (33). For T large and β close to unity we know from (33) that $r=\hat{A}(1+\gamma\alpha)/z$.[11]

VI. *Empirical Tests of the Model*

The model described above has two important and independent features which are readily testable. These two are the technology production function—equation (21)—and the inventor equilibrium—equation (22):

$$\hat{A} = k_0 + k_1N^{\beta}A^{-\gamma} \tag{21}$$

$$\log \dot{N} = k_2 + \left(\frac{1}{1-\beta}\right)\log Q - \left(\frac{\gamma}{1-\beta}\right)\log A + \left(\frac{1}{1-\beta}\right)\log\phi \tag{22}$$

The first equation gives the rate of technological change as a function of the number of inventions and the level of technology. The second gives the flow of inventions as a function of the level of output-technology, and the discount factor, ϕ,

Estimation. It is clear that the theoretical specification gives us a great deal of overidentification since—ignoring disturbances—there are eight coefficients from which only five parameters must be recovered. In fact, we will not be able to use these restrictions in estimation.

In testing the theoretical model we have made two additions. Since the data are from industries, and it is natural to expect that some technological change will come from outside the industries, we have added a constant term, k_0, to equation (21) representing the exogenous rate of technological change. Second, the structure of the random elements must be discussed. It seems unlikely that disturbances in the invention equation (22) will feed back through either Q or A in a period of time sufficiently short as to cause a significant bias. This implies that ordinary least squares estimates of equation (22) will give unbiased estimates. The same cannot be said of equation (21) because there is a lag dependent variable. We must therefore make the stronger assumption that the errors in (21) are serially independent. Once this assumption has been made ordinary least squares estimation will give consistent estimates of the coefficients.[12] Finally, ϕ was dropped from equation (22) because it was statistically insignificant.

---

[11] One of the stock questions asked of such a model is whether it is efficient. The immediate response here would be that it must be inefficient due to the monopolistic restrictions on invention. What then is the optimal legal environment? The optimal kind of legal environment for promoting economic welfare concerns such things as subsidizing or taxing inventive activity and changing the economic or legal life of the invention (say by changing the legal life of the patent). The answer seems to be that in the present model the competitive allocation of resources to invention will be less than is optimal if $T<\infty$. The system can be made efficient either by setting T at infinity or by subsidizing invention. To take a simple numerical example, if $r=.10$, $T=10$, and $(\hat{Q}-\gamma\hat{A})$ is small, then $\phi=6.5$. The social rate of return 0.15. By subsidizing 35 percent of research costs, the system could be made efficient. This conclusion would be modified if labor were not inelastically applied or if there were more than one good.

Equation (22) was estimated by ordinary linear least squares. Estimation of (21) was somewhat more complicated. First round estimates were derived by estimating the equation linearized at its mean. To find the maximum likelihood estimator a grid in the neighborhood of the first round estimates was then examined.

Data. We use the only two industries for which long, continuous time series are available: railroads and agriculture. The most drastic assumption needed is that the annual number of patents issued is a good proxy for the number of inventions. It seems a good bet that the inventions and patents move roughly together. On the other hand, there may well have been a long-run tendency for the propensity to patent to decline. It is, however, hard to see why there would be systematic variations in the propensity to patent over ten- or twenty-year periods. We have used Schmookler's data on patents [4], while output, labor, and capital data are from Kendrick [12]. The rate of total factor productivity was recalculated using a Cobb-Douglas production function with $\alpha=.25$ rather than Kendrick's method. (The two correlate with $\bar{R}^2=.9989$ and $\bar{R}^2=.9997$ for railroads and agriculture, respectively.) The series of ϕ were calculated from series on railroad bond yields from [13] and rates of growth of output.

Results. The empirical results for the two industries are mixed. Roughly speaking, the "technology production function" hypothesis is decisively rejected, while the "inventor equilibrium" hypothesis seems to be consistent with the historical experience.

First examine the production of knowledge hypothesis. There are three different versions. The strictest version of the hypothesis is the nonlinear hypothesis examined in the theoretical analysis above. This equation was estimated by nonlinear methods, as described above, and is reported as regressions 1 and 2 in Table 1. The results are unfavorable to the hypothesis that there is a relationship as specified in the model. Looking at the likelihood ratio of the maximum likelihood estimate compared with the hypothesis that $\beta=\gamma=0$, we find that the hypothesis of a significant difference is rejected at the 50 percent confidence level. Aside from the problem of the statistical significance of the results, at least one of the three coefficients about which one has prior information is wrong in both equations.

Equations (3) through (6) in Table 1 test linear approximations to the technology production hypothesis. In none of these are the results in strong conformity to prior restrictions. Sign patterns are different between the two industries and none of the eight coefficients takes the expected sign at a significant level.

The second part of the hypothesis is the inventor equilibrium. Regressions 7 and 8 in Table 2 give estimates of equation (22), while regressions 9 and 10 introduce a time trend to offset long-run structural changes not encompassed by the model.[13] Examining the equations where a

---

[12] For proof of these statements, see any standard econometrics text. We are in serious trouble if the movements in $\hat{A}$ are due to the business cycle—as they are in the short run. We must hope that any correlation between the business cycle and independent variables in (21) cancel out. On this score we cannot be too optimistic.

[13] The time trend is introduced to pick up systematic changes in the propensity to patent, in the resources needed per invention, and in the cost of inputs.

TABLE 1

ESTIMATES FOR KNOWLEDGE PRODUCTION FUNCTION, 1890–1953*

Number	Industry	Regression Coefficients and *t*-statistics	$\bar{R}^2$	*SEE*	*D−W*
1	Railroad*,†	$\hat{A} = -.0440 + .2809\, N^{-.2} A^{0.}$ $(-.70)\ (1.12)$	.0199	.0594	1.40
2	Agriculture*,‡	$\hat{A} = .221 - .2099\, N^{0.} A^{-.5}$ $(3.10)\ (-2.94)$	.1227	.0543	2.92
3	Railroad	$\hat{A} = .1719 - .0212 \log N - .0105 \log A$ $(1.18)\quad (-.99)\quad (-.43)$	.0216	.0598	1.41
4	Agriculture	$\hat{A} = .0592 + .0105 \log N + .1105 \log A$ $(-.43)\quad (.51)\quad (2.7)$	.1238	.0547	2.90
5	Railroad	$\hat{A} = .439 - .000012\, N - .00201\, A$ $(.99)\quad (-.65)\quad (-.08)$	.0134	.0601	1.40
6	Agriculture	$\hat{A} = -.0808 + .000004\, N + .0868\, A$ $(-.43)\ (.16)\quad (2.46)$	.1141	.0551	2.94

* The minimum was at a corner of the grid, outside the a priori limit given by the theoretical model.

† The approximate standard errors of β and γ were 600. and 50.

‡ The approximate standard errors of β and γ were 60. and 600.

time trend is introduced, the model seems to be appropriate and predicts the signs and magnitudes of the coefficients extremely well. The estimates for β are positive, between zero and one, and not far apart. The coefficients on log A also have the right sign, although that for agriculture is insignificant. We thus have found that the part of the hypothesis predicting inventor behavior performs remarkably well.

As a final test we can use the coefficients estimated and displayed in Table 2 to predict the equilibrium rate of technological change. Using equation (17), we recall that $\hat{A} = \beta\hat{Q}/\gamma$. Although it may involve certain illegitimate assumptions, it is at least a further check on the results to see

TABLE 2

ESTIMATES FOR INVENTOR EQUILIBRIUM

Number	Industry	Regression Coefficients and *t*-Statistics	$\bar{R}^2$	*SEE*	*DW*	$\hat{\gamma}$	$\hat{\beta}$
7	Railroad	$\log N = -1.853 + 1.2867 \log Q - 2.262 \log A$ $(2.72)\quad (12.70)\quad (-19.50)$	.8997	.1878	.632	1.76	.223
8	Agriculture	$\log N = 6.582 + .01288 \log Q - 1.135 \log A$ $(1.96)\quad (.03)\quad (-2.58)$	.3164	.3395	.185	*	*
9	Railroad	$\log N = -2.964 + 1.373 \log Q - 3.026 \log A + .0187\, t$ $(-4.07)\ (13.95)\quad (-11.42)\quad (3.16)$	.9140	.1745	.745	2.20	.272
10	Agriculture	$\log N = -5.678 + 1.896 \log Q - .3245 \log A - .0217\, t$ $(-1.55)\ (3.47)\quad (-.81)\quad (-5.25)$	.5311	.2834	.344	0.17	.473

* $\hat{\beta}$ is outside of the admissible region.

how well it predicts. Table 3 shows the prediction. While they are not too near the actual levels, they at least have the right order of magnitude.

In summary, we have found that the flow of patented inventions behaves according to the predictions of the economic model of the inventive process outlined here. This means, roughly, that patents are positively correlated with increases in output but negatively correlated with the level of productivity. On the other hand, the hypothesis of a relation between the flow of patented inventions and the growth of productivity is not borne out by the two histories examined.

These two results thus imply a significant paradox. While it is easy to reject the technology production function hypothesis because of unfavorable empirical evidence, this leaves the positive results in the other equation—the inventor equilibrium—unexplained. Put differently, if there is no relationship between the flow of new inventions and productivity advance, how can the strong correlations between economic activity and inventions be explained?

Our possibility, of course, is that the historical experience represents a highly improbable occurrence. Thus, if the true β is zero and if the other assumptions (including normality) are correct, the probability that a $\hat{\beta}$ as large as that in either equation is observed is less than five in ten thousand. Clearly an alternative hypothesis would be more congenial.

TABLE 3

PREDICTED AND ACTUAL RATES OF PRODUCTIVITY CHANGE (PERCENT)

Industry	Productivity Change		
	Actual	Prediction I*	Prediction II†
Railroad.	2.59	.35	.45
Agriculture.	1.24	(‡)	2.20

* Using regression 7.

† Using regressions 9 and 10.

‡ Not applicable since estimate was inadmissible.

Rather than throw our lot in with coincidence, the following explanation seems worth consideration. The structure of the random elements may well be such as to lead to the observed results. Several observers have noted that the returns from inventions are highly disperse. (Indeed, this has led many observers to conclude that patent data are worthless as a measure of inventive output.) Scherer [10] reports a rough test showing the value of patents to be distributed according to the arc-sine distribution.

To follow through this line of argument, assume each invention is unpredictable, so that the marginal contribution of the ith invention has a random term ϵ_i which is quite disperse. We can thus rewrite structural equation (21) as

$$(21') \qquad A/A = N^{\beta}A^{-\gamma} + \sum_{i=1}^{N} \epsilon_i(t)$$

Define $u(t) = \sum \epsilon_i(t)$. If the ϵ_i are "well behaved"—say normally distributed—then for given N, $u(t)$ would be distributed with variance $N\sigma^2_\epsilon$. If N is also random, the total variance of u is $N\sigma^2_\epsilon$ if ϵ and N are independent. We might thus expect that any macroeconomic relationship between patents and productivity would be extremely noisy.[14]

On the other hand, if inventors are more or less risk neutral or do not perceive the massive uncertainties of research, then inventor equilibrium in equation (22) will be unchanged. We thus conclude that if the relation between invention and technological change is random, then changes in productivity might be very noisy while the inventor equilibrium might be very well behaved. Both these descriptions fit the empirical results reported here.

---

[14] I am indebted to Zvi Griliches and David Grether for pointing out errors in an earlier version of the paper.

REFERENCES

1. KENNETH J. ARROW, "Economic Welfare and the Allocation of Resources for Invention," in Nelson [8], pp. 609–24.
2. Z. GRILICHES AND D. W. JORGENSON, "The Explanation of Productivity Change," *Rev. of Econ. Studies*, July, 1967, pp. 249–83.
3. JOHN W. KENDRICK, *Productivity Trends in the United States* (Princeton Univ. Press, 1961).
4. JOHN JEWKES, DAVID SAWERS, AND RICHARD STILLERMAN, *The Sources of Invention* (London: Macmillan, 1962).
5. ROBERT E. LUCAS, JR., "Tests of a Capital-Theoretic Model of Technical Change," *Rev. of Econ. Studies*, Sept., 1967, pp. 21–37.
6. FRITZ MACHLUP, "The Supply of Inventors and Inventions," in [8], pp. 143–67.
7. E. MANSFIELD, *Industrial Research and Technological Innovation* (Norton, 1968).
8. R. NELSON, ed., *The Rate and Direction of Inventive Activity* (Princeton Univ. Press, 1962).
9. WILLIAM D. NORDHAUS, *Invention, Growth, and Welfare* (forthcoming from M.I.T. Press).
10. F. M. SCHERER, "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *A.E.R.*, Dec., 1965, pp. 1097–1125.
11. JACOB SCHMOOKLER, *Invention and Economic Growth* (Harvard Univ. Press, 1966).
12. U. S. BUREAU OF THE CENSUS, *Historical Statistics of the United States, Colonial Times to 1957* (Washington, D. C., 1960).

CLASSIFICATORY NOTES ON THE PRODUCTION AND TRANSMISSION OF TECHNOLOGICAL KNOWLEDGE

By KENNETH J. ARROW
Harvard University

I. *Introduction*

Analysis of production functions over the last twelve years has suggested strongly that (*a*) a major proportion of the increase in per capita income cannot be explained by increases in the capital-labor ratio, and (*b*) production functions differ strongly among nations and indeed among regions. Since a production function is defined relative to a given body of technological knowledge, (*a*) implies that technological knowledge has been growing over time and (*b*) that technological knowledge varies over countries.

An economist could just leave the analysis at that, asserting that the causes which determine the amount of technological knowledge at any one time and place lie as much outside his province as the tastes which determine consumption patterns. But in fact, we know that significant quantities of resources are being expended by profit-making institutions on research and development. From the studies of Griliches [4] and Mansfield [8, Part IV], we know that the diffusion of technological knowledge, at least within a given economy, is partly governed by profitability considerations. Hence, it is suggested, we must regard the body of technological knowledge as the result as well as the cause of economic changes.

Economists have had a further, more detailed, preoccupation: that of the bias of technological change. A production function, if it shifts due to increased knowledge, can shift in many ways. In a two-factor model, Hicks spoke of "capital-saving" and "laborsaving" innovations and raised the question whether the bias is itself induced by economic considerations. Fellner [3] convincingly argued that factor prices, per se, should have no tendency to cause bias, since the aim of the entrepreneur is to minimize the total of costs. Kennedy [5] then sought to carry the Fellner analysis further by introducing explicitly the trade-offs between different possible biases in innovation. Kennedy, like most other current writers, has tended to replace the Hicks classification by introducing a more explicit and restricted model of technological change, namely, factor augmentation:

$$Y = F[A(t)\ K, B(t)\ L], \tag{1}$$

where Y is output, K is capital, L is labor, and $A(t)$ and $B(t)$ are the total augmentations of capital and labor, respectively. Then there is postulated a "transformation function" for knowledge, in the form of a trade-off among the rates of growth of A and B and research expenditures.

Models of this type and others (e.g., Arrow [1]) have the natural motivation of using the well-tried tools of production and distribution theory in what appears to be a related field. "Knowledge," as reflected in the variables A and B, appears as an input to physical output, and we then need to supplement the ordinary production relation with an additional relation determining these newly defined inputs.

It is the suggestion of this paper that such models, though they may well be useful descriptively (however, this remains to be tested empirically), do not capture the essential features of the creation and transmission of knowledge. Technological progress is in the first instance the reduction in uncertainty. The product of a research and development effort is an observation on the world which reduces its possible range of variation. The outputs of different research projects are qualitatively different; there is no gain in acquiring the same information twice. The production of knowledge is thus basically different in character from the production of goods, where successive items can be qualitatively identical. The research and development process indeed has quantitative implications for factor productivity, particularly if it is measured at a relatively aggregated level, but these implications are a far from exhaustive description of the research and development process itself.

Research and development is thus intimately connected with the problems of uncertainty reduction which have been the objects of research in mathematical statistics and information theory. The problem of the transmission of knowledge—why different individuals and nations do not have access to equal bodies of knowledge—even more clearly requires analysis of qualitative distinctness of different items of information. Here the disciplines of information and communication theory, learning theory in psychology, and diffusion theory in sociology may be brought to bear.

The remainder of this paper will be devoted to elaborations and exemplifications of these themes. It cannot be claimed that any usable model has yet emerged, but it is hoped that the remarks here will be useful in further developments.

II. *The Production of Knowledge*

Knowledge arises from deliberate seeking, but it also arises from observations incidental on other activities. Haavelmo, Kaldor, and I (see Arrow [1] for references) have all stressed that the activities of production and investment may lead to increases in productivity without any identifiable allocation of resources to that end. An illuminating special case of increased knowledge is the discovery of natural resources. Exploration is analogous to research; but when a community is becoming settled, the members, in the course of their ordinary activities, frequently continue to make unexpected discoveries. Sutter established an agricultural community; his employee, Marshall, saw some yellow flecks in the mill race.

A more general formulation including both research and learning by doing can be formulated and will, I think, be useful. The Bayesian language will be used (for an excellent introduction, see Raiffa [11, Chaps. 1–5]); however, any sensible analysis of uncertainty will lead to equivalent formulations. Let the term "activity" be used, as usual for any process described by inputs and outputs, but we are particularly interested in the case where the outputs are not known with certainty. It can easily happen that the outcomes for different activities will be dependent random variables in the sense of a subjective probability. The case most interesting from the present point of view is that where there is an underlying unknown parameter upon which the probability distributions of outcomes for the different activities depend; then observing the outcome of one activity changes the *a posteriori* distribution of outcomes of the other. This need not happen of course; if the outcome of an activity is known with certainty a priori, then observing the outcome cannot change the probabilities of outcome of any other activity. More generally, if there is statistical independence between the outcomes of activities, the probability of distribution of the outcomes of one is unaffected by observations on the outcomes of another. Thus, if I really believe a pair of dice is fair, observing any outcome is of no use in predicting a subsequent one. But if I suspect bias and I express my suspicions by an appropriate subjective probability distribution over the possible outcomes, then observing an outcome certainly does change my subsequent expectations in accordance with Bayes's theorem.

Thus, an activity will in general have two valuable consequences: the physical outputs themselves and the change in information about other activities.

In many cases, one or the other effect predominates. The classical research situation is one in which the actual output (e.g., of nylon) is of negligible importance compared to the information gain —*a posteriori* the probability that a substance with the properties of nylon can be produced is now 1, whereas a priori it may have been a small figure. In the cases of learning by doing the opposite holds; the motivation for engaging in the activity is the physical output, but there is an additional gain, which may be relatively small, in information which reduces the cost of further production.

Once these polar cases are presented, it becomes clear that intermediate cases are possible, and in fact they not only occur but are, I would hold, frequent. In fact, the bulk of research and development expenditures are actual steps in the production process—design, engineering, tooling, and manufacturing and marketing start-up costs (see Mansfield [7, p. 106]). Each stage involves uncertainties with regard to costs and, at the end, with regard to demand. At each stage, then, something is learned with regard to the probability distribution of outcomes for future repetitions of the activity. At the same time, the physical outputs are expected to be directly valuable.

The problem of the optimal choice of a sequence of activities where both the physical outcome and the information acquired are relevant to the choice is in principle a problem of statistical decision theory, but in fact only fragmentary results have been attained. Analytically, the question is difficult, and, perhaps not surprisingly, the results that have been attained can hardly be stated simply, though definite methods of computing the optimal solutions exist. (In my judgment, this will be an increasingly common situation in economic theory; broad general theorems of the kind we admire can usually only be found under undesirably restrictive conditions. What theory can imply in a broad range of cases is a computing algorithm. To test theory, then, we need econometric evidence or at least well-informed quantitative judgments as inputs to the computing process.) Some beginnings of a microeconomic theory of research and development are found in the work of T. Marschak, Glennan and Nelson (see [10, Chaps. 1, 2 and 5]).

Building a macroeconomic theory on microeconomic foundations is not an easy task; indeed, there is not one completely successful example. But in the case of the production of knowledge, the task is apparently much harder. To proceed from an individual to a collective theory that consumption depends on income or output on inputs involves incompletely justified steps; but it is at least reasonably clear what the aggregates themselves are. Information at the individual level is describable either as the actual outcome of a particular activity or as a whole conditional distribution over states of nature, with the conditions being the actual outcomes. Such a probability distribution is hard to describe in any simple way, and aggregating this information over individuals is even harder.

Given the probabity distribution at any moment of time, individuals make production decisions. The actual outputs are not known with certainty, but one might plausibly take their expected value and regard its relation to the inputs as the production function. This will shift over time as additional information is acquired. The current models in the growth field assume (*a*) that the effect of a given body of information on the production function can be summarized in a few parameters and (*b*) that these same parameters summarize the possibilities for acquiring new information. The first assumption is undoubtedly essential if one is to get anywhere. But the second is on the face of it very misleading. Consider an economy with labor as the only factor of production and operating under constant returns to scale. Then the effect of any given information structure on the production function is completely summarized in the output-labor ratio. But a given output-labor ratio may correspond to very different information structures and therefore very different potentialities for future productivity increases. One might have been based on very thorough exploration of limited areas of investigation, with little further room for increased knowledge, while the other might have investigated more widely and have potential productivity gains available at little additional cost. Securities analysts use such considerations in their evaluations; and historical examples of this difference are frequently cited.

One might be tempted to consider as an alternative measure of aggregate knowledge Shannon's measure of information. There are of course a host of practical difficulties in applying this measure to technological knowledge. But unfortunately it does not seem to be a correct measure for either the supply or the demand side. The irrelevance of Shannon's measure to the demand price for knowledge was shown by J. Marschak [9]. To take an extreme case, consider two states of the world which are indistinguishable with regard to technology; e.g., distinguished only with regard to the number of craters on the dark side of the moon. Any discrimination between these two states would be an increase in information, but surely not one relevant to productivity. On the supply side, similarly, the cost of achieving a given increase in information may differ according to the particular observations made.

III. *Transmission of Knowledge*

The observer of the outcome of an activity can be supposed to form new probability judgments; but how does this affect the information structures of others? The transmission of the observation or of the revised probability judgments must take place over channels which have a limited capacity and are therefore costly. Though the language is borrowed from communications theory, the really limited channels are human minds, not telegraphs or printed words.

Even for the individual there is a problem of channel capacity. To transform his a priori into *a posteriori* probabilities as the result of observations which have taken time, he must remember his a priori probabilities; but memory is a channel for transmission between points of time and is notoriously limited in capacity and subject to error. Natural memory can of course be supplemented with artificial aids —books, files, computer memories—but ultimately there remains a capacity limitation. As a result, even for an individual, the transformation of probabilities and therefore the acquisition of information is less than ideal. Winter in an unpublished paper has argued that memory limitation accounts for the results of the famous Humphreys experiment which has been so crucial in modern learning theory: An event E occurs with probability $p > 1/2$ on each of a succession of trials, with independence between trials. The individual is asked before each trial to predict whether or not E will occur; he is not told the value of p but must in effect infer it from observations. If his aim is to maximize the expected number of successes, the rational strategy would be always to predict E, at least after the process has been observed long enough to establish almost certainly that $p > \frac{1}{2}$. But in fact the subjects wind up predicting sometimes E and sometimes not E, and the frequency with which E is predicted approaches very closely indeed to p. This might be explained by the assumption that they only remember a limited number of previous observations and thus never can be very well assured that it is best to predict E. (To be sure, many other explanations have been advanced in the literature.)

There have been studies by both economists and sociologists on the diffusion of innovations; the relation between the two types of studies resembles nothing so much as the parable of the blind men and the elephant. While Griliches and Mansfield stress the profitability of the investment and the risks involved, the sociologists (see, e.g., Rogers [12] and, for a more theoretical treatment, Coleman [2, Chap. 17]) are concerned with the nature of the channel connecting the adopters of an innovation with potential followers. While mass media play a major role in alerting individuals to the possibility of an innovation, it seems to be personal contact that is most relevant in leading to its adoption. Thus, the diffusion of an innovation becomes a process formally akin to the spread of an infectious disease.

There is nothing irreconcilable in the two viewpoints: in effect, the economists are studying the demand for information by potential innovators and sociologists the problems in the supply of communication channels. Different communication channels have different costs (or, equivalently, different capacities), where these costs include the ability of the sender to "code" the information and the recipient to "decode" it.

The understanding of transmission of knowledge is of especial importance in two of the key socioeconomic problems of our time: (*a*) international inequalities in productivity, and (*b*) the failure of the educational system in reducing income inequality. The two problems have a considerable formal similarity. If one nation or class has the knowledge which enables it to achieve high productivity, why is not the other acquiring that information? That a nation or class has a consistently high productivity implies a successful communication system within the nation or class, so the problem turns on the differential between costs of communication within and between classes.

(It is not intended to assert that difficulties in communication of information are necessarily the sole source of total factor productivity differences. There can be withholding of information to perpetuate monopoly positions; and both in foreign trade and in economic relations between the races, income differentials can arise from exploitation; see Krueger [6]. There might also be genetic factors. But the available evidence certainly suggests that communication problems are a major and perhaps predominant source of productivity and income differentials.)

It may be worthwhile speculating on some of the causes of these differential channel costs. Note first that all the sociological work on diffusion has put great stress on personal contact. Mass media may provide some overall alertness to change, but, except for the most alert and daring innovators, it is the example and advice of those known personally that are apparently most potent in securing acceptance of innovation. As Coleman stresses, however, personal contacts are by no means randomly distributed in the population, and the manner in which any trait (in particular, an item of information) is diffused throughout the system is correspondingly altered. To cite one case, the diffusion of use of new drugs was higher among physicians practicing in pairs than among those practicing singly.

I believe these facts cast light on the basic factors at work in the difficulties of transmission of knowledge across nations and through the educational system. Two general principles may be suggested: (1) A channel has greater capacity if the receiver regards it as more reliable; this is why personal contacts are frequently so important. (2) To a large extent, channels of communication serve purposes other than the diffusion of innovations—friendship, convenience—and the direction of diffusion may be dictated by factors in addition to profitability. Thus international channels, which typically have less other purposes to serve, are more expensive. In particular, and this may be very important, personal contacts across nations are obviously much less than within.

In amplification of (1), there is one special case of unreliability from the receiver's point of view that deserves some attention: the inability of the receiver to understand the message. As already remarked, every piece of information can be regarded as transmitted in a code and can only be used if decoded. In the first instance, a language itself is a code, and the sheer difficulty of translation perhaps can be underestimated. (The inability of English-speaking economists to learn from their French, German, and Italian colleagues is notorious.) There are problems in nonverbal forms of communication. When the British in World War II supplied us with the plans for the jet engine, it took ten months to redraw them to conform to American usage. More subtly, as several gifted observers of the educational scene have observed, there are class and racial differences in the meaning of words, not so much in the literal denotation, but in the connotations and associations, and in the significance of nonverbal behavior. In the complicated interplay of messages between teacher and student, the unreliabilities of communication can lead to extreme inefficiencies.

IV. *Exponential Growth and Stationary Bias*

To conclude, we might briefly consider two aspects of current growth models which are inconsistent with models of technological progress as information-seeking and transmitting.

One is the universal tendency to take some form of rate of growth of productivity as a variable of the system. Typically the model ends up in a quasi-stationary state with a constant rate of productivity growth. Now no known form of adaptive model, with probabilities modified by experience, leads to such a result; on the contrary, such models invariably lead to asymptotes in which output is constant. There is a limit to what can be learned even with infinitely many opportunities. Actually, with respect to the very long run, such a conclusion seems very reasonable. You cannot get something for nothing, ever, and it seems unreasonable to suppose that, by waiting a sufficient length of time, you can get any given output for arbitrarily small input. Eternal exponential technological growth is just as unreasonable as eternal exponential population growth.

But of course exponential technological growth does have the advantage of being consistent with the observed facts; if anything, the observed rate of growth of total factor productivity is increasing. I can only conjecture that, as in the case of population, the true law is something like the logistic curve, but we are still in the early phases, which resemble the exponential.

At the moment, that is, each new item of information opens up the opportunity for acquiring additional items in a constant ratio (in some sense), but eventually the scarcity of additional information will become apparent.

The other aspect is the Kennedy invention-possibility curve, the trade-off between rates of augmentation of different factors for a given research budget. In the first place, there is, of course, no warrant for assuming that technological progress is factor-augmenting; pieces of knowledge about the workings of the world need not be associated in any simple way with different factors of production. In the second place, as already remarked, in the long run, it is unlikely that rates of augmentation are the right variables to enter into any such relation. But most important, there is no warrant for the stability of the curve over time, however it is described. The curve of efficient possible innovations at any time must surely shift with the particular nature of the information acquired, a nature not summed up in productivity measures of any kind. My conjecture is that this is one place where theory should not presume to take the place of history. The set of opportunities for innovation at any one moment are determined by what the physical laws of the world really are and how much has already been learned and is therefore "accidental" from the viewpoint of economics. My guess is that economic factors have little to do with bias in technological progress (though they may have a good deal to do with its magnitude). European desire for spices in the late fifteenth century may have had a good deal to do with motivating Columbus' voyages, but the brute, though unknown, facts of geography determined what in fact was their economic results.

REFERENCES

1. K. J. ARROW, "The Economic Implications of Learning by Doing," *Rev. of Econ. Studies*, June, 1962, pp. 155–73.
2. J. S. COLEMAN, *Introduction to Mathematical Sociology* (Free Press, 1964).
3. W. FELLNER, "Does the Market Direct the Relative Factor-Saving Effects of Technological Progress?" in *The Rate and Direction of Inventive Activity* (Princeton, 1962), pp. 171–88.
4. Z. GRILICHES, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, Oct., 1957, pp. 501–22.
5. C. KENNEDY, "Induced Bias in Innovation and the Theory of Distribution," *Econ. J.*, Sept., 1964, pp. 541–47.
6. A. O. KRUEGER, "The Economics of Discrimination," *J.P.E.*, Oct., 1963, pp. 481–86.
7. E. MANSFIELD, *The Economics of Technological Change* (Norton, 1968).
8. ———, *Industrial Research and Technological Innovation* (Norton, 1968).
9. J. MARSCHAK, "Remarks on the Economics of Information," in *Contributions to Scientific Research in Management* (Western Data Processing Center, Univ. of California, 1960), pp. 79–98.
10. T. MARSCHAK, T. K. GLENNAN, JR., AND R. SUMMERS, *Strategy for R&R* (New York, Springer, 1967).
11. H. RAIFFA, *Decision Analysis* (Addison-Wesley, 1968).
12. E. ROGERS, *Diffusion of Innovations* (Free Press, 1962).

TECHNOLOGICAL PROGRESS AND MICROECONOMIC THEORY*

By WERNER Z. HIRSCH
University of California, Los Angeles

Although this age is characterized by rapid technological progress, economists have been slow to extend the theory of the firm to this phenomenon. This paper attempts to shed light on the nature of technological progress and of the technologically progressive firm. Some telling characteristics of such firms are specified and some special problems faced by economists who seek to analyze them are described. Most of the paper is devoted to discussing aspects of models of technologically progressive firms.

Nature of Technological Progress in the Firm

Technological progress in the firm is the change in its technological knowledge that takes place between two points in time. At any given moment in time a certain body or stock of scientific and technological knowledge has been accumulated by the firm, and provides the firm with a set of techniques; each set is defined as a set of decision and action rules that guide the firm's sequential combination of inputs. Technological progress is evidenced by firms that either produce new or improved products or produce existing products at a lower cost. In an open market economy ideas can be created and exploited for profit; firms will accumulate new scientific and technological knowledge that in turn generates a series of activities designed to translate the new knowledge into new procedures, processes, materials, or products.

How does technology relate to the firm's production function? We can view entrepreneurship, capital goods, workers, and materials as inputs in the production function, with technology its operational part.[1] Technology is described and communicated in a set of instructions for combining inputs so they will result in an output of specified characteristics. Like a cake recipe, technology specifies ingredients as well as sequence of operations. In the neoclassical representation of technological change, improvements in knowledge and techniques result in an alteration of the productivity of inputs; a new production function results.

Here is an example of a new idea that was converted to action rules to guide the sequential application of input factors. After many months of creative deliberations, researchers at Upjohn Laboratories had the idea to use micro-organisms to accomplish molecular restructuring. They devised a set of action rules to put progesterone onto a common mold under specified conditions. One day in late 1950 they placed a Petri dish containing a food substance on a window sill to collect a common mold which was then combined with the sex hormone progesterone. The common mold attached an oxygen atom to the number 11 position of the progesterone molecule, changing it into 11-hydroxprogesterone which is easily processed into synthetic cortisone. Technological progress took the form of a process invention which, in turn, produced a new synthetic product.

Application of new techniques in the firm resulting from the creation of new knowledge is not the sole source of firms' productivity changes. Productivity can

---

* Major assistance in the preparation of this paper was provided by Ronald Teeples and helpful criticism by Donald Shoup and Phillip Vincent.

[1] Richard R. Nelson, Merton J. Peck, and Edward D. Kalachek, *Technology, Economic Growth and Public Policy* (Brookings Institution, 1967), p. 10.

also increase as a result of learning to apply existing techniques more effectively. As time passes, the stock of new scientific and technological knowledge increases; and, as more identical units of a product are produced, opportunities increase for learning to use existing techniques more effectively. Clearly both effects commonly take place at the same time and jointly enhance productivity. The longer the period between the planning of the product and the production of its last unit, and (or) the larger the total contemplated volume of output, the more opportunities the firm has to incorporate new technological knowledge in its production process and to learn to use existing knowledge more effectively.[2]

New technological knowledge in the firm and, therefore, technological progress can result both from internal research and from knowledge transfer. Much new knowledge results from basic research which may be regarded as investigation without specific regard to what the results will be used for; perhaps most new technological knowledge results from applied research—knowledge derived from investigations where the intent is to obtain results for specific uses. New ideas can also be channeled into unexpected and consequently additional uses through transformation of knowledge; i.e., orderly, effective management of the transfer of basic and applied research results to new technological applications.

The Technologically Progressive Firms

Certain firms are in the vanguard of technological progress, creating new products and searching for cost-saving ideas in the hope of staying ahead of their competitors. To initiate steps for developing models of technologically progressive firms we will first specify a few perplexing economic problems that require theoretical insight, and then specify why theoretical models of the types of firms described above would be the most useful approach to solving such economic problems. Here are some examples:

1. Few economists who have investigated models of firms have explored deeply the process by which the policies of firms are arrived at. However, some attempts have been made to explain the behavior of firms by postulating that, because of the separation of ownership and management, they seek to maximize profit subject to a growth constraint. Technologically progressive firms often seem particularly growth-oriented. Is it true that strict owner control over firm policy would imply greater profit orientation and devotion of fewer resources to the systematic development of new products and production processes? Is growth versus profit the most useful theoretical distinction?

2. Why are certain industries characterized by the kinds of firms we call technologically progressive whereas other industries are not? Firm size and industry concentration tend to be the prime correlates examined by economists, but neither of these symptoms seems to offer analytical content to an explanation of why, for example, the drug industry is technologically progressive but the lead industry is not.

3. Does the firm with a large R and D program merely act as a risk-taker, buying up promising projects from independent inventors and consultants and gambling on the projects of educated inventors who cannot find financial intermediaries to finance their endeavors? Or do these firms possess certain economies for

---

[2] Armen Alchian, "Costs and Outputs," in M. Abramovitz, *The Allocation of Economic Resources: Essays in Honor of B. F. Haley* (Stanford Univ. Press, 1959), pp. 23–40; also Kenneth J. Arrow, "The Economic Implications of Learning by Doing," *Rev. of Econ. Studies*, 1962, pp. 155–73.

organizing invention and innovation? Or, are the R and D programs of technologically progressive firms the creation of government; e.g., through government tax provisions, contract requirements, R and D spending, patent laws, or patent enforcement policies? Are certain kinds of invention activities and innovations overproduced and other kinds underproduced because of such policies? Again, models of technologically progressive firms studied under a variety of legal constraints might prove the most fruitful approach to answering these questions.

4. Some portion of the expenditures of technologically progressive firms for advertising can be attributed to the product of information transfer. But what of the expenditures specifically directed at forming new tastes or changing preference intensities? What is the effect of these expenditures on technological progress? What is the optimal level of these taste-affecting activities—for example, of detail men who acquaint physicians with new drugs? Some of this type of advertising is carried out by technologically progressive firms that strive for newness and constant improvement in their product lines. As a result other firms act as if they must counteract their rivals by promoting constant and rapid product change. What is the nature of the good provided by this behavior from the point of view of consumers?

Towards Models of Technologically Progressive Firms

Having identified a few special problems that economists face in analyzing technologically progressive firms, we can now stipulate some important characteristics that should distinguish models of such firms.

Policy Making. The total amount of resources devoted to creating new technological knowledge and the productivity of knowledge-creating activities employed by a firm are direct results of its decision-making process. The neoclassical theory of the firm assumes the decision-maker is a single individual who both owns and manages the firm's resources; the incentives of the decision-maker are to maximize the rate of return to the firm's owned resources; and the decision-making costs are nil except for the direct consequences of decisions that turn out to be unprofitable.

At least two concepts of the decision-making process in neoclassical theory of the firm are found in the literature. The first concept treats the entrepreneur as a technician who, acting as engineer-manager, "employs optimum quantities of inputs" and decides how to "maximize profits" by varying rates of output and prices.[3] And although the formal neoclassical models of the firm do not explicitly show the intertemporal trade-offs, the engineer-manager is assumed to choose the most profitable techniques of production from among all possible production functions. The second concept of the entrepreneur treats him as a resource owner who does not sell the rights to use his resources in production to someone else.[4] Rather, the entrepreneur uses his own resources to produce a product; his reward is the residual between the revenue of the product and its cost. Cost components are the opportunity cost of the entrepreneur's own resources which could have been hired out, and the wages he pays for the cooperating resources he hires by contract. The entrepreneur is viewed primarily as an owner, and technology refers to the amount of knowledge the owner posseses. Technical decisions of production are made with a view to maximizing profits to the owner.

---

[3] James M. Henderson and Richard E. Quandt, *Microeconomic Theory* (McGraw-Hill Book Co., 1958), pp. 42–43.

[4] Milton Friedman, *Price Theory: A Provisional Text* (Aldine Pub. Co., 1962), p. 93.

A different abstraction may be more useful for understanding the behavior of technologically progressive firms—especially the invention and innovation activities carried out by such corporations. Entrepreneurship can be regarded as two separate functions: the first is risk-taking which is ultimately borne by the owner of the firm; the second is technical decision making by the manager of the firm.[5]

The reward to owners of firms—the risk-taking function—is profit defined as capital gains or losses in the value of ownership rights plus the differences between dividend receipts and the riskless opportunity rate of return on the owner's investment. The owner risks his wealth in a given production period subject to the cost constraint of "policing" the employees who are managing the firm. Managers are specifically authorized to make decisions and otherwise use discretion in behalf of the firm's owners.

The reward to manager—the technical decision-making entrepreneurial function—may be analytically segregated into two components: a sanctioned and an unsanctioned income. The sanctioned income is basically related to owner profit. The unsanctioned income may be derived from discretionary uses of the firm's resources and may sacrifice owner profit and hence the manager's sanctioned income. The owner need not be deceived by such uses of the firm's resources but will regard the manager's income as a combination of sanctioned and unsanctioned incomes allowed because owner policing costs are greater than zero.

Thus it does not follow that the firm's investment of retained earnings nor technical decisions, including innovation, will always be carried out to maximize owner profit. Choices of production technique, investments in new physical capital and invention, and the choices involved in innovation may be greatly affected by managers' concern for their own incomes. Differences in R and D expenditures between large and small firms or between private and public firms may be partially explained on the basis of distinguishing risk-taking entrepreneurial functions from technical decision-making entrepreneurial functions and allowing for positive information and enforcement costs.

Production. In the presence of technological progress, the firm must operate with incomplete knowledge about the product it hopes to produce tomorrow and the production method it will employ. Since production techniques frequently change and the repertoires of action relevant to a new technique have to be learned, the technologically progressive firm does not face at any one moment of time a definite and precisely known set of production possibilities. In some respects, production is determined by an evolving historical process, which forces us to abandon the assumption that the individual competive firm can be analyzed with prices treated parametrically and conclusions about market prices deduced at the end.

Matters are further complicated because neoclassical production functions assume contemporaneous transformation of inputs into outputs. They abstract from the economic considerations of multiperiod planning while concentrating on cost effects in a period so short that changes in technological knowledge can be ignored. The conventional long-run planning function, which involves choices of investment in invention, innovation, and imitation, has never been explicitly developed as a choice-theoretic device. Instead the long-run planning function is regarded as merely an envelope for all the one-period functions. The specific benefit-cost choices made by entrepreneurs in moving from one production function to the next are not described but are assumed to be made so the "best" production function is used in each period.

---

[5] A partial development of this notion can be found in Ronald Teeples, "A Working Paper on the Theory of the Firm" (Institute of Government and Public Affairs, U.C.L.A., 1968), OR-143. The basic ideas of enforcement or policing costs appear in Alchian and Allen, *University Economics* (Wadsworth Pub. Co., 1967), second edition. See Chap. 9.

The neoclassical model relates the long-run planning function to the scale of output rather than to time. In this sense the contributions of Armen Alchian and Walter Y. Oi in formulating the long-run planning function with respect to the length of the production process are useful and lend themselves to an analysis of interrelated production functions in successive time periods.[6]

Similarly, the learning function is not compatible with one-period production functions, in which learning, if recognized at all, is regarded as shifts in the production function from one period to the next. Kenneth Arrow and other contributors to learning functions relate technological progress to planning in a way that entrepreneurs gamble on the changes in productivity over time that they believe will result from the experience gained in a more or less specific production process.[7]

Property Rights and Their Transferability. Technological knowledge is created and discovered by man; education, training, and experience are the forms in which technological knowledge is obtained and possessed. Increases in usable technological knowledge are the possession of some individual and are not necessarily an owned asset of the firm. The neoclassical model of the firm regards a change in technological knowledge in one period as an exogenous factor that lowers cost schedules in the following period. But since the firm does not possess all of the property rights, an increase in technological knowledge will not necessarily lead to reduced costs. If increases in the stock of technological knowledge are realized through innovations and (or) experience by employees of a particular firm, the cost schedules will be lower in the following period to the extent that the knowledge is too specific to be of value to other employers (with allowance for the costs of changing jobs, etc.). If knowledge is not specific, the possessors of new knowledge are able to search out higher wage offers; thus gains from the change in productivity due to new technical knowledge in the firm will go to workers and not to the firm.

This problem of changes in technology due to learning can best be handled by a planning function where the technologically progressive firm makes explicit choices between the options of retaining long-term employees who have learned by experience or of instituting training programs for new employees. Furthermore, the technologically progressive firm, aware of its limited ability to enforce property rights related to training and experience provided in the firm, must decide on the mix of general education and highly specific vocational training and professional education it wants to provide its employees.[8] To the extent that any particular activity can be broken down into a set of sub-techniques, each of which requires highly specific training, the firm tends to have relatively firm access to the fruits of its training programs and will tend to emphasize them.

A similar property rights problem exists regarding the firm's adoption of new techniques. The neoclassical model of the firm implicitly assumes that the principal obstructions to the employment of new techniques are the costs of obtaining information, switchover costs, and a reluctance to take risks. However, in the real world there are at least two more important additional constraints: government patent rights over inventions and innovations which prevent the spread of specific technological progress; and government fiat to employees which prevents the adoption of specific techniques of production either by direct law, by interpretation of regulators, or by allowing sufficient union bargaining power to influence firm policy regarding the adoption of those new techniques. Such acts of government influence the bargaining situation between the firm and its employees; and employees' objectives relative to technological progress are often different from the firm's objectives.

---

[6] Armen Alchian, *ibid.;* also Walter Y. Oi, "The Neoclassical Foundations of Progress Functions," *Econ. J.*, Sept., 1967, pp. 579–94.

[7] Kenneth J. Arrow, *ibid.*, pp. 155–73.

[8] Gary Becker, *Human Capital* (N.B.E.R., 1964).

Financing. Firms differ in their capabilities to engage in profitable invention and innovation. A primary difference often singled out as a determinant of technological progress in an industry is the relative ability of a firm to finance invention and innovation. The simplistic argument has been that modern invention and innovation activities are inherently expensive.

The neoclassical model of the firm disregards scale effects in R and D activity as it relates to creation of new products and new techniques of production. Implicit in the notion of a firm's research for technological improvement is the idea that successful R and D increases productivity which in turn gives the firm a cost advantage and thus allows it to enjoy a higher rate of return until the advantage is competed away. Traditionally it was assumed that the sources of successful R and D were the firm's experience in the industry and the fact that it acquired the services of independent inventors. The firm put scientists and engineers on its payroll mainly because the firm's entrepreneur was willing to gamble on the risks of creation and on being able to use the results of R and D. In this scheme, the size of the firm and the scale of R and D activities are irrelevant considerations.

However, scale effects in the R and D process appear to be highly significant—especially for the kinds of product and production technique development that are associated with the technologically progressive firms. The mere fact that significant economies of large scale exist in R and D activities does not automatically grant larger firms a technological advantage over smaller firms; small firms can purchase R and D from large-scale independent R and D suppliers that are organized to exploit large-scale techniques. Thus, although the benefits of large-scale R and D activity need not be limited to large-scale firms, large firms appear to have an advantage in securing economies of large-scale R and D activities, possibly as a result of economies of joint production.[9]

Some of the explanations of the advantages that large firms have for exploiting large-scale R and D may be analyzed by microeconomics. One of the outstanding advantages to large firms is the relatively low cost of capital as viewed by the firm's decision-maker. (Another is the relatively low transaction and communication cost between technical personnel and management, at least up to a certain firm size.) These cost considerations remain to be incorporated into technologically progressive firm models, which must explicitly allow for the determination of internal financing. The models would have to define investable fund sources, a set of possible allocations, relative costs of investable funds, the reward preferences of decision-makers in the firm, and precise constraints on the total amount of investable funds. Such a formulation should explicitly produce the benefit-cost calculation of the decision-maker and thus clear up much of the loose terminology about "firms which require large amounts of investment funds," "firms capable of financing costly R and D," etc.

---

[9] William L. Baldwin, "Contracted Research and the Case for Big Business," *J.P.E.*, June, 1962, pp. 294–98; also, Richard R. Nelson *et al.*, *op. cit.*, especially Chaps. 3 and 4.

Demand. Novel products provided by technologically progressive firms offer new possibilities and choices for users and can affect the development of new tastes and, therefore, demands. The theories of monopolistic competiton, oligopoly, and advertising contribute to an analysis of firms that explicitly allocate expenditures to effect shifts in their demand schedules. However, the interrelation between a firm's demand curves and supply offers has been restricted almost exclusively to an existing known product. Norman Buchanan, among others, has explored how much selling or advertising costs a firm would optimally incur in order to maximize profits from the production of a single, well-defined product.[10]

But we are concerned with the relationship between profits and expenditures of firms that place new products before buyers, because technological progress tends to increase the menu of consumer and producer goods from which a choice can be made. Not merely is a door opened by technology, but we are invited to enter. A firm may offer new options and conditions of choice through technological progress and a firm can alter choice behavior.

Relating a firm's profits to the new and unique products it offers to buyers has two aspects: First, the cost of developing a new product; the firm would like to know the risks and expected costs associated with offering a new product to its buyers. And, second, the firm would want to know the expected demand for the new product at various levels of sales promotion.

A new product, since it tends to compete for sales with existing but not strictly comparable products, should be defined in terms of its demand characteristics rather than in terms of its physical (production) characteristics. A product may be viewed as an entity that gratifies many basic wants or stimuli in the human psyche and can be conceptualized as a vector of "charges" (positive, negative, or zero), one charge for each basic want.[11] In this way a new product may be comparable to, and compete with, existing products that have similar charges for vector elements. Under this definition of a product, the technologically progressive firm that creates new products need not estimate the demand for a good which so far has never been sold. Instead it can estimate the demand for a certain vector of charges, with the relevant market information coming from the sale of existing products in the market that cater to the same abstract desires as the proposed new product.

Resources devoted to the promotion of new products may take two separate forms—informing buyers of the new offer and altering buyers' tastes. Among others, Gordon Tullock applies the vector definition of products to compare the new product with existing ones and recognizes the ability of suppliers to change tastes. However, he concludes that since taste changing efforts are very costly and the effects are unpredictable, taste changing is relatively unimportant.[12] Nevertheless, firms appear to expend much wealth in an effort to change tastes, particularly since these expenditures are not entirely independent of their expenditures to create new products. Possibly, taste changing is less costly if undertaken jointly with the creation of new products. But the fact is that technologically progressive firms in such industries as drugs and electronics appear to play the role of taste-makers as a side product of their efforts to produce new products. Therefore, to understand technologically progressive firms we must explicitly treat the important interdependence between firm supply and demand schedules in the manner described above.

---

[10] Norman Buchanan, "Advertising Expenditures: A Suggested Treatment," *J.P.E.*, Aug., 1942, pp. 537–57.

[11] Kelvin Lancaster, "New Approach to Consumer Theory," *J.P.E.*, April, 1966, p. 132; also, "Change and Innovation in the Technology of Consumption," *A.E.R.*, May, 1966, p. 14.

[12] Gordon Tullock, ***Toward the Mathematics of Politics*** (Univ. of Michigan Press, 1967), see especially Chap. I.

Summary

There is much concern about the behavior of the economy's large corporations, especially about their relation to technological progress. We have attempted to put forth some perplexing economic problems posed by what we call technologically progressive firms and have asserted that the solution of these problems will ultimately depend on producing a theoretic abstraction of large modern corporations.

A fully developed model would allow for the following:

1. The determination of firm policy by more than one actor or set of preferences. This might be accomplished by dividing entrepreneurship into a risk-bearing function and a technical decision-making function.
2. An explicit treatment of the choice of a production technique. The significant feature of this choice is allowance for intertemporal trade-offs in production to establish a long-run planning function.
3. The possession of property rights over changes in technology. Treating the possession of rights either as specific constraints on policy or as factors of production to the firm was suggested in the paper.
4. The determination of the firm's financing. Many allocations of funds to specific inventive and innovative activities would be affected by specific allowance for sources of economies of scale.
5. The determination of the firm's promotion expenditures. By explicitly allowing for taste changing capabilities and the productivity of new products on taste alteration, the paper suggests that an additional aspect of technological progress can be handled.

DISCUSSION

Evsey D. Domar: Let me first discuss the papers by Professors Nordhaus and Arrow and then make a few general remarks.

Papers like Nordhaus' arouse my admiration, if not for their results, then for their ingenuity and courage. If I should register any complaints, it is about an incomplete emancipation of the author from static economics, static not in equations but in spirit. Things like inventor equilibrium (equation 20), a long-run balanced growth path, termination of the inventive process because of the end of population growth, and the like do not really belong to his Schumpeterian world, even if they have become popular in growth models. A negative relation between inventive activity and the optimal stock of capital deduced in his model (and caused by the decline of post-royalty marginal product of capital) also looks strange, as is his assertion that the "movement of all invention from the personal sector to the corporate sector would raise capital's observed share from .18 to .30." Inventive activity has indeed moved toward the corporate sector, but the share of capital has declined.

In his short paper Nordhaus could not tell us how sensitive his results are to changes in some of his assumptions. Suppose that individual inventions were not independent from each other in their effect on productivity, or that some of them had no effect at all unless embodied in capital goods, or that there existed a mutual feedback effect among inventive activities in several industries—which if any of his conclusions would be changed?

If his empirical testing of the relation between inventive activity and the growth of productivity has not been positive, let us not despair and look at it as another courageous attempt to attack a very elusive problem. (Berkev S. Sanders made a similar and unsuccessful attempt a few years earlier; see his "Some Difficulties in Measuring Inventive Activity," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, N.B.E.R., 1962, pp. 53–77.) He may have better luck next time. I am sure he will agree that the number of patents taken need not be a good measure of inventive activity: for a long time this number has been falling in relation to almost any other relevant economic series. Besides, over a period of a decade or two, changes in capacity utilization can easily throw off the regression. (It does not seem that an adjustment for capacity utilization was made.) In the meantime, he should be commended for honestly reporting an unsuccessful test—a most desirable feature in an empirical paper.

That his second test—the relation between economic activity and invention—gave a positive result may be, as he says, a paradox, or it may not be one, since there may be other explanations of this phenomenon, including those suggested by the late Professor Schmookler.

With Professor Arrow's paper I have no quarrels, except for his de-emphasis of the role played by economic factors in what he calls a "bias" in technological progress. It is true that Columbus set out for India and ended up in the Caribbean, but surely against this example one can pile up scores of others on the opposite side. This, however, is a small point. Ever since I built a tree-house for my children I have become a strong adherent of his "learning-by-doing" doctrine, the simplicity of which made me wonder about the process by which ideas are introduced into economic theory (why not analyze the inventive process in our own field?): why economists have been so amazingly late in this respect. Ever since humanity separated itself from the apes, it has been learning by doing, and the performance of my new young dog suggests that animals do likewise, even if less formally. Learning curves have been known in productivity studies for a long time, but somehow this information remained outside the theory of growth. This is not a criticism of Arrow's work; on the contrary, we should be grateful to him for having done recently what should have been done a long time ago. But why pick on learning-by-doing? Couldn't the same be said about technological progress, and the theory of growth in general, and about many other recent achievements based on ideas which have been known to all, including economists, long before their acceptance into the body of economic theory. Perhaps these problems could not be analyzed in a scholarly manner without a good mathematical training not possessed by economists in the past. But is such training required for analyzing the problem of poverty, and particularly poverty amidst plenty, which has bothered humanity at least since biblical times? I wonder how many other unrecognized ideas and problems stare into our faces today.

What these two papers, and for that matter, practically the whole literature on the subject (see for instance the N.B.E.R. volume mentioned above) impress upon us is the strikingly micro character of the process of invention and of technological change. The proper instrument for its analysis, it seems to me, is not aggregate production models like the Cobb-Douglas (which by now does not tell us much more than the inability of the conventionally defined labor and capital to explain a good half of the growth of our output), but the disaggregated model of technological change constructed by W. W. Leontief (*Studies in the Structure of the American Economy,* 1953, pp. 17–52) as early as in 1953 in order to compute what he called the "index of structural change." Surprisingly, little attention has been given to this model, perhaps because fixed input coefficients are not popular among our neoclassical friends. Actually, Leontief's model, when properly aggregated, gives exactly the same result as the Cobb-Douglas (as anyone can ascertain by observing that the rate of growth of the Cobb-Douglas residual equals the weighted mean of the rates of growth of factor productivities, or those of input coefficients with a negative sign). And yet by disaggregating technological change into specific relative changes in pairs of factor productivities (or of their reciprocals) from one period to the next, Leontief's method allows us to observe the economy, if not under a microscope, then under a strong magnifying glass, and to study those changes which an investigator regards important or interesting. It may be possible, for instance, to distinguish active industries which are largely responsible for their own technological change from passive ones (like agriculture or electric utilities) which enjoy the fruits of other industries' efforts.

To an economist who believes in the efficacy of governmental action in promoting technological progress (and particularly to a believer in socialism), Leontief's model imparts a proper sense of humility: technological progress emerges as the result of millions of decisions, each decision contributing a bit, like a bee among the masses in the hive. The model also suggests that although a government can readily increase the rate of capital accumulation and the expenditures on education, research and development, its power to assure that economic decisions are correctly made is rather limited simply because there are so many decisions to be made.

That this is indeed so can be confirmed by even a casual observation of the Soviet experience. In the few areas where the Soviet government has taken a personal interst, so to speak, such as in space, in military hardware and, to a lesser extent, in heavy industry, technological progress has been rapid. But in the rest of the economy, and particularly in the production of consumer goods and services, it has been slow and sporadic. The inventive process as such has not been an obstacle: a large backlog of foreign inventions has always been available. The problem has been in the absence of a built-in mechanism for the introduction of inventions into industry, and particularly of inventions resulting in better and more modern products. If anything, the system has developed an automatic device for the rejection of such inventions, or at least for a long delay in their acceptance. And ironically enough, this has been caused, at least in part, by Arrow's learning-by-doing.

Let me explain. The learning-by-doing process enables an enterprise manager to master a given activity and to reduce the cost. Thus an old product is both easy and profitable to produce since its price, once established, remains fixed for a long time. A new product, or a new model, naturally involves all sorts of difficulties and uncertainties, and the price for it is fixed at the outset with but a small profit margin. I doubt that the recent Soviet shift from output targets to profit maximization (more or less) is likely to make a radical change in this attitude. Hence, the paradox that when learning-by-doing process is combined with a system of inflexible and arbitrary prices, it may retard rather than accelerate the rate of technological progress.

JOSEPH STIGLITZ: These three papers begin with the premise—with which I strongly concur—that to further our understanding of the increases in productivity of recent decades, we need to examine microeconomic behavior. From here, however, the three authors take off in quite different directions.

The first difference that strikes one is the level of "disaggregation" to which they turn. Nordhaus modifies the familiar neoclassical growth model (with its representative firms) to explain how an economy, which is competitive in the product market but with temporary monopolies in inventions, allocates savings between research and innovation; Hirsch focuses on the individual firm; and Arrow on the process of production and transmission of knowledge itself.

A second important difference is that of the two characteristics of research which are commonly asserted to make it distinct from other economic activities—the uncertainty associated with the outcome, and the difficulties associated with appropriating the returns (the public goods aspects)—Arrow focuses on the former, Nordhaus on the latter, Hirsch discusses both.

This is all for the good: all of these diverse approaches are undoubtedly needed. But somewhat disturbing is that while one of Arrow's conclusions denies the existence of any stable relationship between inputs into research and "research outputs," Nordhaus' model is predicated on the existence of just such a relationship (and the absence of uncertainty). Can these views be reconciled? An economist's immediate response is to appeal to aggregation, which may reduce the uncertainty (but does not necessarily make the relationship stable from period to period), but Arrow cautions us that aggregation is not likely to work.

Consider a one final good, two-factor economy. In Arrow's model, the producer has a distribution function over the output of that commodity for different inputs into the production activities. Inputs into "research activities" change the distribution function. Arrow's assertion is that, consequently, the "output" of a particular research activity cannot be summarized in a single number, and in this he is undoubtedly correct.

But there may be a systematic relation between the input of factors into research activities and changes in the average value of output per unit input of factors in production. It is these increases in productivity which provide the motivation for the research, and even its measure of economic value. Indeed, some econometricians, e.g., Griliches and Mansfield, have claimed to have found such stable relations. It should be noted that Arrow is willing to use the average value of output per unit input in his production function.

Of course, a macroeconomic relationship between inputs into research and changes in productivity may have little relevance to microeconomic analysis, for firms, in allocating resources to research may (in fact, are likely to) take cognizance of the uncertainty associated with the outcome from any particular research project. (The same objection could be raised to the production function if there is uncertainty present.)

Arrow suggests that the problem of allocating resources to research is sufficiently complicated that analytic results are simply not to be had; that advances will come only through the numerical solution of particular problems, employing the techniques of statistical decision-making theory. It appears to me that it is premature to give up on the analytic approach; that in fact there are unresolved conceptual problems in the application of statistical decision theory which mean that we are not yet ready to deliver this problem to the computer. For instance, what is the firm's utility function, the existence of which is typically posited by theories of statistical decision making, and how do competitive asset markets relate it to the utility functions of individuals?

Whether or not economists and operations researchers can eventually convince firms that this is how they ought to allocate their resources, this approach leaves the economist without answers to some of the more important questions which they seek to answer; e.g., how ought our patent system to be organized? For answers to these questions, the economist must, I think, turn to the kinds of models that Professors Nordhaus and Hirsch have been examining.

The second aspect of research, the "public goods property," arises from the fact that the transmission cost of the results of research are much lower than the cost of research itself; accordingly, unless artificial (government) barriers are placed on the free transmission the firms doing the research will not be able to capture (appropriate) the returns, and too little research will be undertaken. Yet those barriers themselves will cause a distortion between price and marginal costs. In Professor Nordhaus' model, no loss of welfare arises from the difference between price and marginal cost because all factor supplies (and commodity demands) are perfectly inelastic. If, however, there are no perfectly inelastic demand curves and commodities are produced by constant returns to scale production functions, then some loss of welfare arises from the attempt to raise revenue to pay for the research. This must be balanced off against the benefits of the research.

If we view expenditure on research as a public good, it is clear that, if lump-sum taxes do not exist, the presence of the divergence between price and marginal cost does not in itself imply that the economy has failed to attain a (second best) Pareto optimum. If the supply elasticity for capital (savings) is identical to that for labor, in a one-good model the situation described by Professor Nordhaus would in fact be a (constrained) Pareto optimum. The question arises, however, of whether, in general, the patent system is as good as some alternative fiscal system with respect to (*a*) the total allocation of resources to research, (*b*) the allocation of research among alternative industries, (*c*) the total dead-weight loss resulting from the divergence between price and marginal cost. The patent system has two peculiar characteristics: (*a*) it is like a benefit tax, in that the consumers of the products of each industry are required to pay for the research in that industry; (*b*) the "tax" (divergence between price and marginal cost) is at a fixed level (equal to the savings induced by the invention) for one period, and zero for the remainder of the life of the invention. Let us compare the patent system with a system of excise taxes and research subsidies (in a competitive model). Before we can do that, however, we must determine what we mean by competitive equilibrium in the research sector. If we assume, with Nordhaus, that the research production function is strictly concave, we could have an equilibrium with total profits zero or marginal profits zero. Nordhaus prefers the latter. This makes sense in a single period model where there are a number of firms already in the research industry, who have already "taken their profits." An additional firm would then only get the "marginal profits." If, however, the firms are really identical (as they appear to be in Nordhaus' model), it is hard to distinguish between the marginal and intramarginal firms: they should all be receiving the same profits. Free entry will lead then to total (average) profits being zero.

Let us assume, now, that we have two small industries each with a well-defined research function and the government is constrained to having taxes equal subsidies for the two industries together but not each separately.[1] We know then that (constrained) Pareto optimality will require the ratio of the percentage divergence between price and marginal cost to be (approximately) equal to the inverse ratio of the elasticities of demand.[2] Notice this has nothing to do with the research potentialities of the two industries, only the demand functions. The optimal allocation to research, on the other hand, requires that the present discounted value of cost savings from an additional unit of research (at the levels of demand resulting from the correctly chosen taxes and subsidies) equal the marginal cost of an additional unit of research. Such a (constrained) Pareto optimum could be achieved either by the government taking over the research industry and collecting the royalties at the appropriate rate, or by the government letting patents have an infinite life, but taxing the royalties (in the case of the zero profit equilibrium) or imposing a (100 percent) tax on profits in the research industry and with the revenues subsidizing the production of the commodities.

---

[1] The following remarks are based on the simple model described in footnotes 3 and 4 below.

[2] See F. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.*, Mar., 1927, and M. Boiteux, "Sur la Gestion des Monopoles Publics astreints à l'equilibre budgétaire," *Econometrica*, Jan., 1956.

Now let us impose the additional constraint that each industry's tax revenues must equal its subsidies. First order maximum conditions, analogous to those of the previous problem, can then be written down, although now, of course, we consider each industry separately.[3] Let us consider the effect of the constraint making each industry pay for itself on the allocation of a given research budget: if the two industries have identical research and production functions, the constraint will result in more research being done in the industry with the relatively inelastic demand function (than if no such constraint exists); if the two industries have identical demand and production functions (before research begins), then the constraint will reduce the input of research into the industry with the relatively elastic research function (the one for which a percentage change in research inputs results in the greater percentage cost reduction). In both cases, since the cost of obtaining revenue for financing the research is greater under the constraint than without the constraint, and the benefits gained from the research are less, the optimal level of research will be less with the constraint than without it.

The mechanism for attaining the equilibrium described above is identical to that of

---

[3] To see this, consider the following simple model. The two commodities, C_1 and C_2 are produced by labor alone; the wage is expected to be constant at unity, and the interest rate at r. The demand functions are assumed to be static, $C_i = d_i(p_i)$. Research lowers the labor requirements per unit of output according to the functions $\alpha_i(L_{Ri})$. With the usual caveats about the use of consumers' surplus, we can view the first problem as finding the solution to

$$\max_{\{p_i^*, L_{Ri}\}} \sum_i \int_0^\infty \int_{p^*}^{p_{i0}} d_i(p_i) e^{-rt} dp_i dt = \gamma \sum \int_{p_i^*}^{p_{i0}} d(p_i), \qquad \gamma = \frac{1}{r} \tag{1}$$

subject to the constraint

$$\gamma \sum_i [p_i^* - p_{0i}\alpha_i(L_{Ri})] d_i(p^*_i) = \sum_i L_{Ri}, \tag{2}$$

where p_{0i} is the price (cost) before research, p_i^* the price after research. The first order conditions are (we shall assume here and below that there is a unique, interior solution):

$$\frac{\tau_1/p_1^*}{\tau_2/p_2^*} = \frac{1/n_1}{1/n_2} \tag{3}$$

where $\tau_i = p_i^* - p_{0i}\alpha_i(L_{Ri})$ (price minus marginal cost), $n_i = |d_i' p_i/d_i|$—the ratio of the percentage deviations of price from marginal cost should equal the reciprocal of the ratio of the elasticities of demand—and

$$\gamma p_{01}\alpha_1' C_1 = \gamma p_{02}\alpha_2' C_2 = -1. \tag{4}$$

The ratio of revenues in the two industries is

$$\frac{\tau_1 C_1}{\tau_2 C_2} = \frac{p_1^* C_1/n_1}{p_2^* C_2/n_2}$$

so that the ratio of the ratio of expenditures on research to revenues is (using (4)):

$$\delta = \frac{n_1 m_1 p_{01}\alpha_1}{p_1^*} \Big/ \frac{n_2 m_2 p_{02}\alpha_2}{p_2^*}$$

where $m_i = -\alpha_i' L_{Ri}/\alpha_i$, the elasticity of the research function. If $\delta < 1$ there is a net transfer of revenues from the first to the second sector. If the two industries have identical research functions (say of constant elasticity), and $p_{01} = p_{02}$, then

$$\delta - 1 = (n_1 - n_2)/(n_2 p_{02}\alpha_2/p_2^*) \gtrless 0 \quad \text{as } n_1 \gtrless n_2$$

If the two industries have identical demand functions (say of constant elasticity), and $p_{01} = p_{02}$, then $\delta = m_1/m_2$. The second problem, where each industry must pay for itself, is

$$\max \gamma \int_{p_i^*}^{p_{0i}} d_i(p_i) \tag{5}$$

subject to the constraint

$$\gamma [p_i^* - p_{0i}\alpha_i(L_{Ri})] \, d_i(p_i^*) = L_{Ri} \tag{6}$$

(4) still holds, but instead of (3) we now have

$$\frac{\tau_1/p_1^*}{\tau_2/p_2^*} = \frac{m_1/(1+\gamma m_1)}{m_2/(1+\gamma m_2)} \tag{7}$$

Observe that (3) involves only demand elasticities, (7) only elasticities of the research function.

the previous case: either the government does the research directly itself, or it issues a patent with an infinite lifetime, taxes the profits, and subsidizes the production of the commodity.

Let us now look at the patent system, without subsidies or taxes. If we issue an infinite lifetime patent, consumers gain nothing from the research activity, and, in the zero average profit competitive equilibrium, neither do producers or inventors. With no patent, there is no research, and this clearly is not optimal. An optimal patent lifetime is one which maximizes the present discounted value of the consumer surplus generated by the research over the period after the patent has expired (since with zero average profit competitive equilibrium, there is no producer surplus; the modification for the other case is straightforward). If we compare the previous case with our optimal patent situation, we will find that in the former situation (i.e., with taxes) consumers are generally better off, research in general will be higher, percentage of the total and dead-weight loss as a potential consumer surplus (i.e., if price were set equal to marginal cost) in general is smaller. (These properties are always true if the demand curves are approximately linear or concave and the research function has approximately constant elasticity.)[4]

If now we ask how the patent system distorts the allocation of a given research budget between different industries, we obtain results similar to that obtained earlier. If we now require, in addition, that the lifetime of the patent be the same in both industries, then the misallocation of the research becomes even larger. Whereas the correct allocation requires the marginal savings from an extra unit of research be the same in both industries, the patent system will equate the average savings per unit of research.[5]

---

[4] More formally, the optimal patent lifetime (using the simple model of the previous footnote) is derived as follows for the zero profit competitive equilibrium. (Here, as elsewhere, we assume that each invention is relatively small, and consequently does not confer upon the inventor any monopoly power. For a further discussion of optimal patents, see W. D. Nordhaus, "The Optimal Life of a Patent," Cowles Foundation Discussion Paper 241, 1967.)

$$(8)\quad \text{Max}\int_{T}^{\infty}\int_{p_0\alpha(L_R)}^{p_0}[d(p)e^{-rt}dp]dt = (1-\beta)\gamma\int_{p_0\alpha(L_R)}^{p_0}d(p),\quad 0\leq\beta=1-e^{-rT}\leq 1$$

Subject to:

$$(9)\quad C_0\int_0^T p_0(1-\alpha(L_R)e^{-rt}dt = p_0\beta\gamma(1-\alpha(L_R)C_0 = L_R$$

where $C_0 = d(p_0)$. The first order conditions for this are:

$$(10)\quad p_0(1-\beta)\gamma d(p_0\alpha(L_R))\alpha' + \mu[\beta\gamma\alpha' p_0C_0 + 1] = 0$$

and

$$(11)\quad \int_{p_0\alpha(L_R)}^{p_0} d(p) = \mu(1-\alpha(L_R))p_0C_0$$

where μ is the Lagrangian multiplier associated with the constraint (9). To see the three stated propositions, observe that if we denote the optimal β by β^*, and L_R by L_R^*, and if, in the second problem (equations 5 and 6) we set $\tau/p_0(1-\alpha)=\beta^*$, $L_R=L_R^*$, $\tau d(p_0\alpha(L_R)+\tau)>L_R^*$, so that we can make $\tau<\beta^*$ and still obtain the requisite revenue.

But then the consumer surplus is greater than under the optimal patent. (*A fortiori*, if we choose (τ, L_R) optimally in the second problem, consumer surplus will be larger.) From (10) and (11) we have

$$(12)\quad \frac{[(1-\beta^*)d(p_0\alpha(L_R)) + (\mu\beta^*)d(p_0)]\gamma p_0|\alpha'|}{\mu} = 1$$

where $\mu=(\overline{C}/C_0)>1$, where

$$\overline{C} = \int_{p_0\alpha(L_R)}^{p_0} d(p)/p_0(1-\alpha(L_R))$$

If d is approximately linear (or concave), the left-hand side of (12) is less than

$$\gamma|\alpha'|p_0d(1-\beta)p_0\,\alpha(L_R)+\beta p_0),$$

so at L_R^*, $\gamma|\alpha'|p_0d(p_0\alpha(L_R)+\tau)>1$, so it pays to undertake still further research. Using (9), (12), and (6) we obtain, if α has constant elasticity,

$$\frac{\beta}{\hat{\tau}/(1-\alpha)p_0} \geq \frac{(1-\beta^*)d(p_0\alpha(L_R))+\beta^*\overline{C}}{\overline{C}} > 1.$$

Thus, if the demand curves are approximately linear (or concave), then the ratio of the dead-weight loss to the total potential consumer surplus is greater under the optimal excise tax than under the optimal patent.

[5] Evaluated at $C_0=d(p_0)$. From (9) if $\beta(T)$ is the same for both industries:

$$\frac{p_{01}\gamma(1-\alpha(L_{R1}))C_{01}}{L} = \frac{1}{\beta} = \frac{p_{02}\gamma(1-\alpha(L_{R2}))C_{02}}{L_{R2}}.$$

RESEARCH AND DEVELOPMENT AND OTHER DETERMINANTS OF INVESTMENT

INVESTMENT AND THE FRUSTRATIONS OF ECONOMETRICIANS*

By ROBERT EISNER
Northwestern University

"Always study your residuals," Paul Samuelson has advised "the scientific forecaster."[1] But the econometrician working in the field of investment might well scream, "Which residuals?"

For indeed the first question is what relation to estimate. And immediately beside that problem, perhaps literally simultaneous with it, is the question of what measured variables to take as proxies for the variables of our theoretical relation. At the root of the difficulty is the fact that investment, more even than other variables, is forward looking, dependent upon a relation between initial conditions, which we may hopefully more or less know, and expectations of the future about which both the investment decision-maker himself and his trailing econometrician are frequently singularly ill-informed. In our econometric models the investment relation includes variables such as current and past output, sales, profits, stock yields, interest rates, depreciation charges, stocks of capital, age of capital, capacity, prices of output, prices of labor and of capital goods, "liquidity," a host of tax parameters, composite variables measuring "cost of capital" and "rental price of capital," a great variety of lags, many dummy variables relating to specific factors, behavioral units, sub-aggregates, or time periods, and some measures of expectations.

Investment functions have been estimated from data at the level of national aggregates, industries, firms, and establishments. Investigators have variously used cross-sections and time series and "overall" or hybrid relations involving observations for cross-sections of units such as firms or industries over different time periods.

The underlying theory from which we might hope to derive a relation to estimate would not seem all that difficult. Essentially we think of firms trying to maximize their present values, given a set of capitalization rates of expected future returns and given a set of initial conditions which includes conspicuously the existing stock of capital. With output constrained by a production function, investment is derived directly from the solution path for capital stock, taking into account the depreciation or wearing out of capital implied by the solution output and composition and amounts of capital called for by the production function.

Complications begin to develop rapidly, however. Recognition that competition is not perfect forces the etching in of so many extra lines on the broad canvas as to make the picture seem to many almost unrecognizable. For one thing, obviously enough, for sets of expected future prices we must now substitute sets of expected future demand curves with possibly varying elasticities over time and with elasticities even subject to control or influence by the firms themselves. The downward sloping demand curves do at least have the virtue of enabling our theory of investment to bring in demand considerations (as opposed to purely price parameters) at the level of the individual firm as well as at the level of the industry or the economy as a whole. But as we also recognize less than perfect elasticities of supply, we add the seriously complicating, although clearly realistic, dimension of cost of adjustment. The optimum or solution capital stock at any point of time will almost certainly be different as we recognize that the costs per unit of all factors of production are related to the rates at which they are being acquired and, very likely, the relations between rates of acquisition at one point of time and rates at other points of time.[2]

---

* Among recent contributors to the continuing research cited in this paper I should single out Betty Benson, who continues to help me improve and complicate my computer program, and my able research assistant, Allan Mendelowitz. I am particularly indebted to Charles W. Bischoff for the quarterly data and some of the formulations reported upon in Table 6, and to Dale W. Jorgenson for the data used again in Table 7. For the main body of individual firm data I once more thank the McGraw-Hill Publishing Company and Margaret K. Matulis of McGraw-Hill. Grateful acknowledgement is made of the financial support of the National Science Foundation and a Ford Foundation Faculty Fellowship and a partially concurrent term as a Fellow at the Center for Advanced Study in the Behavioral Sciences.

[1] [14], p. 129.

The rational (and informed) "entrepreneur," then, must plan a path for capital over time which maximizes his firm's present value, taking into account not only all of the initial conditions and current parameters—production function, supply and demand functions, and tax structure—but the expected values of these parameters over all of the relevant future. What is worse, each decision-maker is in the usual oligopolistic quandry, attempting to estimate his optimal moves in full recognition that these depend upon the responses of other firms to the moves that he makes or that the other firms expect him to make. And perhaps complicating matters still further, it is not entirely reasonable to assume that firms in a world of risk and uncertainty act as if they are trying to maximize their present values. They may rather wish to maximize a function in which the mathematical expectation of present values enters positively while the variance of the probability distribution of anticipated present value outcomes enters negatively. And indeed in a world of uncertainty and imperfect information the decisions which maximize current market value may not be those that entrepreneurs will rationally make if they are interested in maximizing some function of wealth over time. An entrepreneur may wisely decide upon an investment path which he is confident will raise the value of his firm in the future by reducing risk and hence lowering capitalization rates, or otherwise favorably affecting the value of the firm in manners not currently anticipated by the market.

These complications may begin to reveal the difficulty of the task faced by the econometrician. Operating at the level of the individual firm, where can he even begin to get information as to the relevant variables? He hardly has any notion of the production function or the demand and supply functions of the past or present, let alone those anticipated for the future. This information is barely available to the business decision-maker himself, and then in a scattered or desultory form which may prove operationally meaningful to him but is scarcely the stuff for consistent quantification in econometric models.

What then do we do? We simplify! Now simplification is obviously of the essence of scientific procedure but it is important that the simplification preserve the essentials of the relationship with which we are concerned, and it is not at all clear that in much of our work with investment this is so. For the first, overwhelming simplification involves substituting past or current variables for those anticipated future variables which are relevant. We are bound by the availability of data and we simply do not have data as to what demand, supply, relative prices, production functions and tax structures will be in the future or even at all precisely what economic behavioral units expect them to be in the future. And then, since we hardly have reliable information on any of the functions we have been talking about, we usually deal with observations, for the past and present, of points on these functions. Thus we find ourselves dealing not with demand functions but with past sales, not with supply functions but with current and past rates of factor acquisition.

---

[2] See [13] for pathbreaking empirical estimates of simultaneous adjustment in a multifactor model.

There are methods of rationalizing this. We might, for example, argue that shifts in demand are expected to be iso-elastic and marginal costs constant so that prices of output will remain unchanged. We might then assume that expected future output will be some given or estimable function of past and current sales or output. We might assume that the elasticity of all price expectations is unity, meaning that all changes in current prices, or some weighted average of current and past prices, will be reflected in proportionate changes in expected future prices.

In some cases, present variables, because of limitations on the perfection of markets, may actually serve well. If firms' costs of capital are determined at the time capital expenditures begin and financing is arranged, part of the issue of expected future prices may be dismissed. But capitalization rates may yet change in the future, altering the present value of expected future returns and affecting the actions of later competitors whose future investment will then influence the value of investment currently being undertaken.

With all of these pitfalls in the path of research on investment, it is indeed easy to throw stones at those who have fallen in the quest of truth. I am not normally reticent in pointing out the (real or imagined) limitations of other investigations, but perhaps it would be fairest to begin, if not to concentrate, on some of the difficulties in my own work. For as I proceed with analysis of the mass of McGraw-Hill capital expenditure survey and related data on which I have reported on a number of occasions, some findings fall more and more solidly into place, but others stand out as a warning to the unwary and a tribute to the intractability of our problem.

By now, working with capital expenditure surveys from 1955 through 1966, I am able to report the results of regressions involving as many as 3,702 observations of individual firms in ten broad industry groups over the twelve-year period. In one relation we have assumed capital expenditures as a ratio of gross fixed assets to be a function of current and six lagged annual relative sales changes, the ratio of profits to gross fixed assets, current and lagged, and the ratio of depreciation charges to gross fixed assets. The sales change variables may be taken as a proxy for changes in expected future demand, the depreciation variable as a measure of replacement requirements (or is it flow of funds?), and the profits variables for everything else left out which might affect the expected profitability of investment. And in some ways the results, as shown for example in Table 1 for the "firm overall" regression, are encouraging and enlightening. As one would expect from good acceleration principles (or a production function with diminishing marginal products to each factor but positive cross-partial derivatives) the sales change coefficients are all clearly significantly positive and sum to a substantial .486, indicating that a 10 percent change in sales (demand or output?) will result in a corresponding 5 percent change in capital stock over a six-year period. Noting the means of each of the variables, the sales change coefficients imply that the elimination of growth of sales, *ceteris paribus,* would on the average reduce capital expenditures of the firm by some 25 percent. A similar perusal of the other coefficients and means suggests that, with no profits, capital expenditures would be some 11 percent less while with no depreciation charges (and replacement requirements?) capital expenditures would be 44 percent less. Finally, we are left with the rather disconcerting significantly positive constant term of .019, arguing that, with zero values of all of our independent variables, capital expenditures equal to 1.9 percent of 1957 gross fixed assets, some 20 percent of total capital expenditures, would none the less be made.

One can spend a lot of time interpreting, rationalizing, and correcting these findings—and on some future occasion I shall undoubtedly do so—but at the moment it might be wisest to remind the patient reader that this set of results is only one among many. The same observations can be grouped to obtain pooled cross-sections within industries and across industries or cross-sections of firm means (where the means of the observations of each firm, for all of the years for which its observations are included, become the observations, appropriately weighted, of our regression). And we can also work with industry time series, industry cross-sections and industry overall regressions, in which the observations in each case are the means for the firms within the industry-year-group.[8] Further, we may report on regressions without the rather mischievous depreciation term. The results, I must quickly confess, are reasonably different in the different regressions and I have indeed elsewhere [4] tried to capitalize on these differences by relating them to some meaningful hypotheses

TABLE 1

McGraw-Hill Data. Capital Expenditures, 1955–66, as a Function of Sales Changes, Profits and Depreciation*

$$i_t = b_{10} + \sum_{j=1}^{7} b_{1j}\Delta s_{t+1-j} + \sum_{j=8}^{9} b_{1j}p_{t+8-j} + b_{1.10}d_{53} + u_1$$

Firm Overall Regression

Variable or Statistic (1)	Regression Coefficients and Standard Errors (2)	Means and Standard Deviations and Products (3)
Constant term or i_t	.019 (.002)	.093 (.077)
Δs_t	.080 (.009)	.062 (.128)
Δs_{t-1}	.102 (.009)	.052 (.126)
Δs_{t-2}	.077 (.009)	.046 (.125)
Δs_{t-3}	.072 (.009)	.040 (.121)
Δs_{t-4}	.070 (.009)	.041 (.119)
Δs_{t-5}	.051 (.009)	.042 (.118)
Δs_{t-6}	.034 (.009)	.036 (.121)
p_t	−.021 (.026)	.106 (.104)
p_{t-1}	.126 (.027)	.101 (.096)
d_{53}	.775 (.040)	.053 (.028)
$\Sigma\Delta s$ Coefficients and, in Col. (3), Σb_{1j} (Mean Δs_{t+1-j})	.486 (.023)	.023
Σp Coefficients and, in Col. (3), Σb_{1j} (Mean p_{t+8-j})	.106 (.012)	.011
$n(-249)$†	3702	
r.d.f.‡	3691	
$\hat{R}^2$§	.302	

* Variables listed in Tables 1 through 5 are as defined in [4], p. 370. Thus, $i_t = I_t/K_{57}$, $\Delta s_t = 3(S_t - S_{t-1})/(S_{56}+S_{57}+S_{58})$, $p_t = P_t/K_{57}$, and $d_{53} = D_{53}/K_{53}$, where I = gross capital expenditures, S = sales, P = profits after taxes, D = depreciation charges, and K = gross fixed assets. Classification of observations, by firm and industry, and statement of deviations used in various regressions are to be found in [4], pp. 388–89.

† n = number of observations; the figure in parentheses on this line is the number of individual firm observations eliminated because of extreme values for one of the variables. (See [4], pp. 369–70.)

‡ r.d.f. = residual degrees of freedom.

§ $\hat{R}^2$ = adjusted or unbiased coefficient of determination.

---

[8] See [4], pp. 388–89, for definitions of the deviations used in the various regressions.

about permanent and transitory variances. Some of the differences are displayed, for the latest available data, in Table 2. We find now that the sum of sales change coefficients in regressions including the depreciation term varies from .380 to .553 in individual firm observations and from .466 to .732 where the observations are industry-year-means.

The differences become most marked, as seen in Table 3, when we compare time series, cross-section and overall regressions without the depreciation term (which had to be eliminated from the time series because, taken as of 1953, it did not vary over time). In the firm time series, it may be noted, the sum of the sales change coefficients was only .294, as against a very significantly different .717 in the cross-section of firm means. Looking further at this latter regression, we find that it implies that some 34 percent of capital expenditures is associated with growth in sales and some 14 percent with the existence of profits, while fully 52 percent is impounded in a constant term which may, as ever, reflect errors in the variables we have included—or the effects of myriad variables, including this time any reasonable measure of replacement requirements, which we have excluded. And in the industry cross-section the sum of sales change coefficients mounts to .911, not significantly different from unity, while the sum of the profits coefficients is only .123, not significantly different (at the .05 probability level) from zero. This relates some 46 percent of gross capital expenditures to growth in sales and some 14 percent to profits. The remaining 40 percent stems presumably from replacement requirements or some form of change in product or process.

But we have many more results to mention, if now only fleetingly. To permit more meaningful comparisons over time, and for some other reasons, we have estimated relations in which capital expenditures and profits were deflated by gross fixed assets of the previous year and sales changes were deflated by an average of sales of the current year and two previous years. The sum of sales change coefficients is now somewhat less (.355) in the firm overall regression, perhaps in part due to the larger proportion of more recent observations included and the lower coefficients for cross-section regressions of later years, possibly related to lower variance of the capital expenditure variable, as defined in these regressions, for later years.

Pursuing the matter further in Table 4, it becomes clear that there is not intertemporal stability in the cross-sectional relations. The F-ratio for reduction in residuals by using individual year cross-sections instead of the pooled regression for all years is 2.57, pointing to statistically significant heterogeneity. (The .01 probability level of F for the indicated degrees of freedom is 1.43.) What is more, the simple arithmetic mean of sums of sales coefficients for the six years of regressions involving capital expenditures of 1955 to 1960 is .360, while the corresponding mean for 1961 to 1966 is only .184. May this relate to the fact that the mean relative sales increase was only 4.5 percent from 1955 to 1960, and only 3.8 percent for the six years before that, as against 7.0 percent from 1961 to 1966? Perhaps when demand is generally growing the differences in the growth of demand among firms (which were in fact less in the later years) have less to do with differences in capital expenditures. Perhaps! But this does not square too easily with Meyer and Kuh's suggestion that the accelerator is more evident in expansionary periods,[4] or my own earlier findings

---

[4] [12], p. 134.

TABLE 2

Capital Expenditures, 1955–66, as a Function of Sales Changes, Profits and Depreciation*

$$i_t = b_{20} + \sum_{j=1}^{7} b_{2j}\Delta s_{t+1-j} + \sum_{j=8}^{9} b_{2j} p_{t+8-j} + b_{2,10} d_{63} + u_2$$

Firm and Industry Cross-sections and Industry Overall Regressions
Regression Coefficients and Standard Errors

Variable or Statistic (1)	Firm Cross-section				Industry	
	Within Industries (2)	Across Industries (3)	Means within Industries (4)	Means Across Industries (5)	Cross-section (6)	Overall (7)
Constant Term	.029# (.003)	.021# (.002)	.027# (.006)	.021# (.006)	.013# (.007)	.009 (.007)
Δs_t	.067 (.009)	.068 (.010)	.106 (.053)	.060 (.057)	.012 (.076)	.075 (.064)
Δs_{t-1}	.063 (.009)	.076 (.009)	.045 (.062)	.100 (.067)	.173 (.064)	.247 (.041)
Δs_{t-2}	.057 (.009)	.067 (.009)	.054 (.060)	.059 (.066)	.123 (.059)	.151 (.038)
Δs_{t-3}	.056 (.009)	.062 (.009)	.096 (.066)	.122 (.071)	.065 (.056)	.114 (.041)
Δs_{t-4}	.060 (.010)	.061 (.010)	.060 (.072)	.031 (.078)	.044 (.059)	.083 (.046)
Δs_{t-5}	.051 (.010)	.053 (.009)	.053 (.069)	.057 (.075)	.036 (.055)	.034 (.046)
Δs_{t-6}	.027 (.009)	.030 (.009)	.120 (.054)	.122 (.058)	.012 (.056)	.028 (.040)
p_t	−.023 (.025)	−.025 (.026)	−.007 (.129)	.042 (.142)	.207 (.249)	.114 (.232)
p_{t-1}	.155 (.026)	.120 (.027)	.111 (.139)	.017 (.152)	−.342 (.265)	−.187 (.247)
d_{63}	.633 (.047)	.828 (.039)	.594 (.081)	.783 (.072)	1.344 (.160)	1.069 (.150)
$\Sigma\Delta s$ Coefficients	.380 (.025)	.418 (.024)	.535 (.054)	.553 (.054)	.466 (.121)	.732 (.106)
Σp Coefficients	.132 (.012)	.095 (.012)	.104 (.024)	.059 (.024)	−.135 (.061)	−.073 (.061)
$n(-249)$†	3702	3702	503	503	119	119
r.d.f.‡	3573	3680	483	492	97	108
$\bar{R}^2$§	.187	.274	.348	.452	.628	.662

*, †, ‡, §. See footnotes for Table 1.
# Composite.

TABLE 3

Capital Expenditures, 1955–66, as a Function of Sales Changes and Profits*

$$i_t = b_{30} + \sum_{j=1}^{7} b_{3j}\Delta s_{t+1-j} + \sum_{j=8}^{9} b_{3j}p_{t+8-j} + u_3$$

Firm and Industry Time Series and Cross-sections and Firm Overall Regressions

Variable or Statistic (1)	Firm Time Series (2)	Cross-section of Firm Means (3)	Firm Over-all (4)	Cross-section of Means within Industries (5)	Industry Time Series (6)	Firm Cross-section within Industries (7)	Industry Cross-section (8)	Firm Cross-section Across Industries (9)	[(3)−(2)] Cross-section of Firm Means Minus Firm Time Series (10)	[(8)−(7)] Industry Cross-section Minus Firm Cross-section within Industries (11)
Constant Term	.045∤ (.002)	.048 (.006)	.051 (.002)	.053∤ (.006)	.014∤ (.010)	.058∤ (.002)	.037∤ (.008)	.053∤ (.002)	.004 (.007)	−.021 (.008)
Δs_t	.063 (.009)	.130 (.063)	.090 (.010)	.119 (.056)	.052 (.048)	.070 (.010)	.115 (.099)	.081 (.010)	.067 (.064)	.045 (.099)
Δs_{t-1}	.068 (.008)	.058 (.075)	.103 (.009)	.024 (.065)	.131 (.039)	.064 (.009)	.258 (.083)	.085 (.010)	−.010 (.075)	.194 (.083)
Δs_{t-2}	.046 (.009)	.068 (.074)	.086 (.009)	.067 (.064)	.097 (.034)	.061 (.009)	.190 (.077)	.078 (.010)	.022 (.074)	.129 (.077)
Δs_{t-3}	.040 (.009)	.152 (.079)	.086 (.009)	.109 (.069)	.100 (.036)	.063 (.010)	.124 (.073)	.078 (.010)	.112 (.080)	.061 (.073)
Δs_{t-4}	.041 (.009)	.006 (.087)	.083 (.010)	.043 (.076)	.113 (.037)	.067 (.010)	.083 (.076)	.077 (.010)	−.035 (.088)	.016 (.077)
Δs_{t-5}	.026 (.009)	.107 (.083)	.070 (.010)	.081 (.072)	.086 (.036)	.060 (.010)	.093 (.072)	.071 (.010)	.081 (.084)	.033 (.073)
Δs_{t-6}	.010 (.008)	.196 (.064)	.052 (.009)	.171 (.056)	.069 (.030)	.037 (.010)	.047 (.073)	.048 (.010)	.186 (.065)	.011 (.074)
p_t	.095 (.025)	.016 (.158)	.000 (.027)	−.003 (.136)	.154 (.169)	−.012 (.025)	.156 (.325)	−.001 (.027)	−.079 (.160)	.168 (.326)
p_{t-1}	.242 (.026)	.111 (.169)	.162 (.029)	.136 (.146)	.331 (.196)	.170 (.027)	−.033 (.344)	.159 (.029)	−.131 (.171)	−.204 (.345)
$\Sigma\Delta s$ Coeffic.	.294 (.033)	.717 (.058)	.569 (.024)	.614 (.056)	.649 (.132)	.423 (.025)	.911 (.142)	.518 (.025)	.423 (.067)	.488 (.144)
Σp Coeffic.	.338 (.025)	.127 (.026)	.162 (.012)	.133 (.025)	.485 (.133)	.158 (.012)	.123 (.068)	.158 (.012)	−.210 (.036)	−.036 (.069)
$n(-249)$† $r.d.f.$‡	3677 3165	503 493	3677 3667	503 484	119 100	3702 3574	119 98	3702 3681	$F(9; 3658)$ $= 10.20$	$F(9; 3672)$ $= 10.44$
$\hat{R}^2$§	.160	.321	.230	.277	.633	.145	.363	.187	$F_{.01} = 4.32$	

*, †, ‡, §. See footnotes to Table 1.

∤ Composite.

TABLE 4

CAPITAL EXPENDITURES, 1955–66, AS A FUNCTION OF SALES CHANGES, PROFITS, AND DEPRECIATION. ALL EXCEPT DEPRECIATION MEASURED AS RATIOS OF PREVIOUS YEAR'S GROSS FIXED ASSETS OR PREVIOUS THREE YEAR SALES AVERAGE.*** INDIVIDUAL YEAR AND POOLED FIRM CROSS-SECTIONS**

$$i_t^* = b_{40} + \sum_{j=1}^{7} b_{4j}\Delta s^*_{t+1-j} + \sum_{j=8}^{9} b_{4j} p^*_{t+8-j} + b_{4,10} d_{53} + u_4$$

Variable or Statistic	Years												All, Pooled	Mean 1955–60	Mean 1961–66
	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966			
	Means and Standard Deviations														
i_t^*	.095 (.078)	.111 (.075)	.105 (.077)	.074 (.058)	.074 (.068)	.075 (.062)	.064 (.055)	.061 (.047)	.062 (.042)	.071 (.043)	.083 (.056)	.093 (.075)	.080 (.064)	.089	.072
Δs_t^*	.099 (.122)	.067 (.115)	.035 (.111)	−.043 (.116)	.087 (.110)	.024 (.101)	.026 (.110)	.079 (.090)	.067 (.086)	.083 (.091)	.083 (.088)	.084 (.098)	.056 (.110)	.045	.070
	Regression Coefficients and Their Sums and Standard Errors														
Year Dummies from Pooled Regressions	.034 (.005)	.047 (.004)	.044 (.004)	.023 (.004)	.018 (.004)	.020 (.004)	.012 (.004)	.007 (.004)	.005 (.004)	.011 (.004)	.018 (.004)	.026 (.004)			
Constant Term	.007 (.012)	.024 (.008)	.037 (.008)	.033 (.007)	.019 (.008)	.036 (.007)	.028 (.006)	.016 (.005)	.027 (.005)	.033 (.006)	.033 (.008)	.025 (.010)	.021# (.002)	.026	.027
Δs^*_t	.088 (.041)	.165 (.036)	.144 (.041)	.106 (.038)	.011 (.037)	.111 (.036)	.063 (.027)	−.019 (.027)	.025 (.027)	.021 (.029)	.038 (.044)	.124 (.050)	.078 (.010)	.104	.042
$\Sigma\Delta s^*$ Coeffic.	.306 (.104)	.470 (.069)	.423 (.076)	.320 (.062)	.339 (.073)	.305 (.077)	.262 (.066)	.210 (.059)	.092 (.060)	.230 (.062)	.187 (.087)	.121 (.114)	.322 (.021)	.360	.184
Σp^* Coeffic.	.115 (.047)	.158 (.035)	.011 (.043)	−.035 (.036)	−.065 (.041)	.048 (.040)	.057 (.039)	.118 (.031)	.206 (.033)	.116 (.035)	.106 (.053)	.124 (.073)	.080 (.012)	.039	.121
d_{53}	1.221 (.205)	.828 (.144)	.980 (.155)	.911 (.123)	1.121 (.128)	.560 (.123)	.503 (.106)	.541 (.086)	.326 (.086)	.332 (.090)	.561 (.118)	.925 (.161)	.667 (.035)	.937	.531
$n(-181)$†	189 (−11)	262 (−14)	267 (−16)	239 (−13)	304 (−24)	322 (−22)	324 (−19)	304 (−12)	264 (−14)	264 (−10)	236 (−12)	255 (−14)	3,230 (−181)	264 (−17)	274 (−14)
$r.d.f.$‡	178	251	256	228	293	311	313	293	253	253	225	244	3,208	253	264
$\hat{R}^2$§	.371	.464	.365	.369	.320	.158	.196	.272	.237	.200	.186	.190	.254††	.341	.214

**, †, ‡, §. See footnotes to Table 1.

# Composite.

*** Variables denoted by asterisk in Tables 4 and 5 are deflated by K_{t-1} (in the cases of capital expenditures and profits) and $(S_t+S_{t-1}+S_{t-2})/3$ (in the case of sales changes), rather than K_{57} and $(S_{58}+S_{57}+S_{56})/3$, respectively, as in Tables 1, 2, and 3. Further explanation is to be found in [4], especially p. 370.

†† $F(110; 3{,}098) = 2.57$; $F_{.01} = 1.43$.

that sales change coefficients are higher in firms whose sales are growing or growing more rapidly.[5]

We may also examine cross-section results by broad industry-group, and here again, in Table 5, despite normalization of the variables so that relations will not reflect directly interindustry differences in capital-output ratios, we find substantial and significant heterogeneity. Sales change coefficients sum to .436 in the growing, strongly demand-influenced utilities, but profits coefficients are also high here. Can this really be a flow-of-funds relation among predominantly large firms which raise the bulk of their capital through borrowing? The other industry with large sales change coefficients is, unaccountably, primary metals, with a record of relatively modest growth in sales (an average annual increase of only 2.36 percent). The *F*-ratio on differences among regressions is 1.87, again well above the .01 probability level value of 1.47. The pooling of firm time series by individual industries interestingly offers some confirmation of the cross-section results by industry, in that utilities and primary metals again reveal the highest sums of sales change coefficients, but various of the profits coefficients, including those in utilities, are surprisingly high. And once more we have significant interindustry differences.

There are many other tables I could present, were time and space available. Some would show uncertain results with the use of current and lagged depreciation variables, others some role for a variable measuring changes in the market value of the firm, and still others a positive contribution, as might be expected, for the ratio of actual to desired utilization of capacity. And we can also show that capital expenditure anticipations are relatively good forecasters of actual expenditures and that sales expectations as well as actual sales are positively related to inventory investment. But throughout there remain enigmas, not the least of which are the differences in results from the various groupings of individual firm, industry and aggregate cross-sections, time series and overall regressions.

This work has little to say about the role of relative prices, nor has it as yet been formulated to attempt to explore the determinants of a possibly variable speed of adjustment. But there is other work that tries to deal with these matters, with a number of critical assumptions, the crucial nature of which may all too soon be forgotten in the understandable enthusiasm to present "significant" results. I have just seen one recent paper—as yet unpublished and unquotable—which essays to measure the roles of accelerated depreciation and the investment tax credit by specifying an investment relation to which are added dummy variables for the years in which accelerated depreciation and the investment credit each were operative. From the coefficients of these dummy variables come estimates, to the millions of dollars, of the effects of these major fiscal measures. Can any of us be at all confident that the dummy variables so defined are capturing the effects of the specified fiscal measures rather than any of the rest of the virtually infinite potential influences on capital expenditures from 1954 to 1961 and 1962 on?

And yet, how much better off are we when we presume to specify the role of our fiscal measures in variables and relations which reflect as well so many other unspecified factors? I have been involved in some controversy over the interpretation to be made of provocative recent reports by Jorgenson and his associates[6] as to the role of relative prices and the contribution of various tax incentives to investment as measured by their specified relation to relative prices. With Nadiri I have elsewhere shown[7] that appropriately unconstrained estimation of a "neoclassical" relation using Jorgenson's own data suggests a vastly smaller role for relative prices—perhaps none at all—than that assumed by Jorgenson in his specification of a Cobb-Douglas production function. But one may well argue that the fault is to be found somehow in Jorgenson's data rather than in his hypothesis and, while objecting to the assumption of unitary or any other specified elasticity of the demand for capital stock with respect to relative prices, we should admit that its true role is yet to be determined. One may therefore welcome additional work on this subject currently being undertaken by Charles Bischoff and, vulture that I am with other people's data, I am able to report briefly on the results of further estimation of Jorgenson-type, neoclassical relations (but with elasticities unconstrained) using the Bischoff data. These data are restricted to equipment but cover all of the private sector of the U.S. economy, unlike the data taken from Jorgenson, which included equipment and structures but involved only manufacturing. The Bischoff data have the advantage of extending through 1966, so that some more adequate test of the role of the investment tax credit and acceleration of depreciation in connection with the new "guidelines" may be examined. The results are intriguing, even if they raise as many questions as we may hope they will settle.

---

[5] [3], pp. 16, 22.

[6] [8] [9] [10] and [11], among others.

First, working with a relation precisely analogous to that which I used with Nadiri, and restricting ourselves to the quarterly data through 1962, we get a very direct confirmation of what was discovered previously with the Jorgenson data. Our best estimate (in terms of minimizing the squares of the residuals) indicates an elasticity of capital stock with respect to output of .68 and of capital stock with respect to relative prices (which according to Jorgenson is + 1.0) a nonsignificant −.08. When we take the data through 1966 our elasticity of capital stock with respect to output rises to .85 and that with respect to relative prices to .05, as seen in the top half of Table 6.

Perhaps I should have quit when I was ahead! But alas, Bischoff has himself led in another path and candor requires me to follow it as well, and to report my results, which in part confirm his. Working with the independent variables defined as logarithms rather than differences in logarithms, our elasticity of capital stock with respect to output for the period through 1962 rises to .98 and that with respect to relative prices to a quite respectable .58! And when we operate through the 1966 period, the elasticity with respect to output goes to 1.03 and that with respect to relative prices to .78! Is Jorgenson's hypothesis reasonably confirmed after all, however unreasonable it may be to take it on faith without testing?

My answer is to suggest that we remain agnostic. Bischoff sees the better fit and absence of evidence of serial correlation in the residuals of the relations in the straight logarithms, in some cases, as an argument that those relations are superior. But rejection of the hypothesis of serial correlation in the disturbances would appear uncertain in relations such as these, with lagged values of the dependent variable. Also, with the Jorgenson data in the one relation in straight logarithms (with $s = 1$) where extreme multi-collinearity does not make it impossible to estimate (with $s = 2$ my computer reports a negative residual variance!), the low estimate

---

[7] [6]; see also [5].

TABLE 5

CAPITAL EXPENDITURES, 1955–66, AS A FUNCTION OF SALES CHANGES, PROFITS, AND DEPRECIATION, ALL EXCEPT DEPRECIATION MEASURED AS RATIOS OF PREVIOUS YEAR'S GROSS FIXED ASSETS OR PREVIOUS THREE YEAR SALES AVERAGE:**
CROSS-SECTIONS AND TIME SERIES BY INDUSTRY*

Variable or Statistic (1)	Primary Metals (2)	Metal Working (3)	Chemical Processing (4)	All Other Manufac-turing (5)	Mining (6)	Utilities (7)	Petroleum (8)	Railroads (9)	Stores (10)	Transportation and Communication (11)	All Industries (12)
Means and Standard Deviations											
i_t^*	.065 (.060)	.087 (.064)	.075 (.047)	.074 (.050)	.074 (.055)	.066 (.029)	.068 (.037)	.029 (.016)	.118 (.087)	.189 (.123)	.080 (.064)
Δs_t^*	.038 (.125)	.078 (.145)	.064 (.106)	.045 (.087)	.047 (.129)	.049 (.060)	.061 (.078)	.009 (.096)	.053 (.082)	.098 (.090)	.056 (.110)
p_t^*	.054 (.041)	.122 (.098)	.093 (.081)	.124 (.105)	.099 (.075)	.033 (.006)	.064 (.070)	.021 (.016)	.132 (.103)	.037 (.033)	.089 (.089)
d_{63}	.047 (.022)	.068 (.027)	.053 (.017)	.051 (.016)	.054 (.026)	.022 (.004)	.048 (.020)	.018 (.004)	.065 (.026)	.113 (.039)	.052 (.029)
Regression Coefficients and Their Sums and Standard Errors Cross-sections											
Δs_t^*	.129 (.034)	.069 (.016)	.064 (.021)	.088 (.026)	.023 (.057)	.076 (.021)	.067 (.033)	.046 (.015)	.126 (.067)	.255 (.137)	.075 (.010)
Σs^* Coefficients	.505 (.082)	.244 (.040)	.239 (.045)	.192 (.057)	.028 (.130)	.436 (.048)	.110 (.068)	.221 (.059)	.413 (.118)	.385 (.324)	.265 (.022)
Σp^* Coefficients	.146 (.096)	.136 (.023)	.075 (.025)	.101 (.022)	−.074 (.120)	.702 (.196)	.089 (.036)	.185 (.066)	.177 (.053)	.800 (.391)	.114 (.012)
d_{63}	.400 (.138)	.455 (.073)	.171 (.112)	.506 (.141)	.221 (.258)	−.079 (.300)	.646 (.115)	.917 (.246)	.493 (.205)	.713 (.282)	.451 (.043)
$n(-181)$†	251 (−13)	768 (−73)	521 (−14)	468 (−16)	83 (−14)	402 (−13)	153 (−4)	216 (−5)	261 (−26)	107 (−3)	3230 (−181)
$r.d.f.$‡	229	746	499	446	61	381	131	194	239	85	3101
$\hat{R}^2$§	.225	.202	.110	.145	−.082	.227	.244	.164	.143	.127	.145

$F(90; 3{,}011) = 1.87$ $F_{.01} = 1.47$

TABLE 5—(*Continued*)

Variable or Statistic (1)	Primary Metals (2)	Metal Working (3)	Chemical Processing (4)	All Other Manufacturing (5)	Mining (6)	Utilities (7)	Petroleum (8)	Railroads (9)	Stores (10)	Transportation and Communication (11)	All Industries (12)
						Time Series					
Δs_t^*	.151 (.035)	.061 (.015)	.037 (.020)	.056 (.027)	−.041 (.049)	.012 (.020)	.006 (.027)	.059 (.012)	.066 (.053)	.228 (.166)	.051 (.009)
$\Sigma\Delta s^*$ Coefficients	.436 (.124)	.228 (.049)	.141 (.072)	−.098 (.095)	−.213 (.164)	.408 (.061)	−.072 (.104)	.268 (.054)	.215 (.159)	−.299 (.643)	−.219 (.031)
Σp^* Coefficients	.805 (.131)	.415 (.035)	.501 (.057)	.336 (.047)	.281 (.129)	1.523 (.309)	1.275 (.149)	.412 (.113)	.414 (.150)	2.112 (.653)	.447 (.024)
n	251	757	515	454	83	402	150	216	257	107	3192
$r.d.f.$‡	212	637	437	376	61	350	120	181	204	85	2744
$\hat{R}^2$§	.369	.303	.192	.135	.053	.246	.406	.322	.078	.209	.177
			$F(81; 2{,}663)=3.62$			$F_{.01}=1.50$					

*, †, ‡, §. See footnotes to Table 1.
** See footnote to Table 4.

TABLE 6

BISCHOFF QUARTERLY DATA FOR EQUIPMENT.* ESTIMATES OF ELASTICITY OF CAPITAL STOCK WITH RESPECT TO RELATIVE PRICES ($\hat{E}_p$) AND OUTPUT ($\hat{E}_q$), FROM DIFFERENCES IN LOGARITHMS (6.1) AND FROM STRAIGHT LOGARITHMS (6.2), FOR 1949–62 AND 1949–66

$$\Delta \ln K_t = \alpha_1 + \sum_{i=1}^{7} [\gamma_{1vi} \Delta \ln V_{t-i} + \gamma_{1qi} \Delta \ln Q_{t-i}] + \sum_{j=1}^{s} \omega_{1j} \Delta \ln K_{t-j} + u_1 \tag{6.1}$$

$$\Delta \ln K_t = \alpha_2 + \sum_{i=1}^{7} [\gamma_{2vi} \ln V_{t-i} + \gamma_{2qi} \ln Q_{t-i}] + \sum_{j=1}^{s} \omega_{2j} \ln K_{t-j} + u_2 \tag{6.2}$$

$$\hat{E}_p = \sum_{i=1}^{7} \hat{\gamma}_{vi} \Big/ \left(b - \sum_{j=1}^{s} \hat{\omega}_j\right) \qquad \hat{E}_q = \sum_{i=1}^{7} \hat{\gamma}_{qi} \Big/ \left(b - \sum_{j=1}^{s} \hat{\omega}_j\right)$$

$$b = 1 \text{ in } (6.1), \qquad b = 0 \text{ in } (6.2)$$

Relation (1)	Data Through Year (2)	s (3)	$\Sigma\hat{\gamma}_{vi}$ (4)	$\Sigma\hat{\gamma}_{qi}$ (5)	$\Sigma\hat{\omega}_j$ (6)	$\hat{R}^2$ and SEE (7)**	$\hat{E}_p$ (8)	$\hat{E}_q$ (9)	$b-\Sigma(\hat{\gamma}_v+\hat{\omega})$ (10)†	$b-\Sigma(\hat{\gamma}_q+\hat{\omega})$ (11)†	ρ and $D-W$ (12)‡
(6.1)	1962	2	−.0221 (.0327)	.1946 (.0547)	.7154 (.0581)	.8838 .001800	−.0776	.6836	.3067 (.0685)	.0900 (.0549)	−.1274 2.25
	1962	1	−.0259 (.0331)	.1633 (.0515)	.7271 (.0585)	.8800 .001829	−.0948	.5985	.2988 (.0694)	.1096 (.0542)	−.2086 2.41
(6.1)	1966	2	.0115 (.0246)	.2071 (.0511)	.7576 (.0534)	.8895 .001806	.0476	.8453	.2309 (.0602)	.0353 (.0487)	−.0042 1.97
	1966	1	.0088 (.0245)	.1869 (.0475)	.7615 (.0534)	.8892 .001808	.0367	.7835	.2297 (.0603)	.0516 (.0463)	−.0754 2.12
(6.2)	1962	2	.0251 (.0106)	.0422 (.0234)	−.0430 (.0218)	.9190 .001503	.5827	.9813	.0179 (.0237)	.0008 (.0035)	−.2189 2.40
	1962	1	.0398 (.0102)	.0780 (.0218)	−.0796 (.0196)	.9033 .001642	.5000	.9798	.0398 (.0246)	.0016 (.0039)	.4118 1.10
(6.2)	1966	2	.0318 (.0064)	.0421 (.0194)	−.0407 (.0182)	.9308 .001429	.7809	1.0349	.0089 (.0200)	−.0014 (.0017)	−.1083 2.17
	1966	1	.0489 (.0052)	.0746 (.0196)	−.0735 (.0181)	.9136 .001597	.6661	1.0158	.0245 (.0219)	−.0012 (.0019)	.1945 1.49

* $V = e^{\lambda t}(p/c)$, where p is an index of the price of output, c is a measure of the rental price of capital involving the industrial bond yield and the dividend-price ratio, and λ is taken as .00178 from nonlinear estimation of an investment function as reported by Bischoff in [1, p. 66] and [2, p. 10]. Q is constant dollar gross product of the business sector.

** $\hat{R}^2$ = adjusted or unbiased coefficient of determination; SEE = standard error of estimate.

† The values of the statistics in columns (10) and (11) would be zero if the corresponding elasticities, $\hat{E}_p$ and $\hat{E}_q$, were unity.

‡ ρ = serial regression coefficient $= \Sigma u_t u_{t-1} / \Sigma u_{t-1}^2$. $D-W$ = Durbin-Watson statistic.

of $\hat{E}_p$ is confirmed, as seen in Table 7, and the relation in differences in logarithms gives a better fit.

It would seem to me that there must be something systematic at work in the Bischoff data to produce estimates so different for two relations, that of the logarithms and that of the differences in logarithms, which are *theoretically* equivalent. Are there two or three broad swings in the

TABLE 7

JORGENSON QUARTERLY DATA FOR MANUFACTURING.** ESTIMATES OF ELASTICITY OF CAPITAL STOCK WITH RESPECT TO RELATIVE PRICES ($\hat{E}_p$) AND OUTPUT ($\hat{E}_p$), FROM DIFFERENCES IN LOGARITHMS (7.1) AND FROM STRAIGHT LOGARITHMS (7.2), FOR 1949–62

$$\Delta \ln K_t = \alpha_1 + \sum_{i=1}^{7} [\gamma_{1pi} \Delta \ln (p/c)_{t-i} + \gamma_{1qi} \Delta \ln Q_{t-i}] + \sum_{j=1}^{s} \omega_{1j} \Delta \ln K_{t-j} + u_1 \tag{7.1}$$

$$\Delta \ln K_t = \alpha_2 + \sum_{i=1}^{7} [\gamma_{2pi} \ln (p/c)_{t-i} + \gamma_{2qi} \ln Q_{t-i}] + \sum_{j=1}^{s} \omega_{2j} \ln K_{t-j} + u_2 \tag{7.2}$$

$$\hat{E}_p = \sum_{i=1}^{7} \hat{\gamma}_{pi} \Bigg/ \left(b - \sum_{j=1}^{s} \hat{\omega}_j\right) \qquad \hat{E}_q = \sum_{i=1}^{7} \hat{\gamma}_{qi} \Bigg/ \left(b - \sum_{j=1}^{s} \hat{\omega}_j\right)$$

$$b = 1 \text{ in } (7.1), \qquad b = 0 \text{ in } (7.2)$$

Relation (1)	s (2)	$\Sigma\hat{\gamma}_{pi}$ (3)	$\Sigma\hat{\gamma}_{qi}$ (4)	$\Sigma\hat{\omega}_j$ (5)	$\hat{R}^2$ and SEE (6)*	$\hat{E}_p$ (7)	$\hat{E}_q$ (8)	$b-\Sigma(\hat{\gamma}_p+\hat{\omega})$ (9)†	$b-\Sigma(\hat{\gamma}_q+\hat{\omega})$ (10)†	ρ and $D-W$ (11)‡
(7.1)	2	.0066 (.0121)	.0751 (.0276)	.8742 (.0439)	.9308 .001279	.0525	.5974	.1192 (.0472)	.0506 (.0423)	−.0388 2.00
	1	.0058 (.0122)	.1040 (.0159)	.8865 (.0438)	.9297 .001289	.0432	.7786	.1278 (.0471)	.0296 (.0403)	.0931 1.75
(7.2)	2			Too much multicollinearity to estimate. ($r_{\ln K_{t-1}, \ln K_{t-2}}$=.99924; R^2 estimated as 1.00968!)						
	1	.0080 (.0048)	.1197 (.0095)	−.1579 (.0120)	.8315 .001996	.0508	.7581	.1498 (.0126)	.0382 (.0039)	.7146 .62

*, †, ‡. See footnotes to Table 6.

** K=net capital stock, p= the wholesale price index, c=a rental price of capital based on the U.S. government long-term bond rate, and Q=gross value added, deflated by p. Data are described further in [6] and [10].

stocks or flows which dominate the logarithmic relation and wash out in the first difference form? Does the relation in logarithms which finds a substantial price elasticity reflect any more than the equivalent of a dummy variable for the investment boom in the 1960's? Does V, involving dividend-price ratios, reflect largely stock market swings and expectations of profit on investment, as pointed out in [6], rather than relative prices?

The vector of simple correlations of $\Delta ln\ K_t$ with $ln\ V_{t-i}$, $i = 1$ to 7, we may note, is [.691, .747, .794, .848, .850, .831, .779], while that of $\Delta\ ln\ K_t$ with $ln\ Q_{t-i}$ is [.001, −.022, −.054, −.088, −.113, −.128, −.137]. It may be added that all of the elements of the correlation matrix involving the $ln\ K$ and $ln\ Q$ variables are upwards of .98, which would seem to guarantee $\hat{E}_q$ in the neighborhood of unity, and leave the autocorrelated $ln\ V$ variables to pick up whatever broad swing in $\Delta\ ln\ K_t$ is left "unexplained." Do we really have 72 independent observations or two or three, reflecting each of the major movements of the postwar period? Are these broad swings then really associated with the variables we have specified or with innumerable others, unspecified, as well as all of the expectational ramifications of those we have included? Indeed, it occurs to me in this election year, might we have fared better by adding the logarithm of the Democratic Party vote?[8]

[8] One may also wonder whether a systematic relation with the "dependent variable, $ln\ K_{t-1}$ . . . the natural logarithm of a constant dollar measure of the

And so we come back, full circle. It is a tricky business, designing a precise investment relation from *ex post* variables which fit more or less remotely the usually expectational variables of our true investment relation. It seems highly probable that the relations in the variables which we measure are unstable. How and how soon business firms will react to changes in relative prices, sales, the rate of invention, and changing information, to mention just a few factors, are matters receiving and demanding much additional study. Sadly, perhaps, it would seem to remain as proper now as it was almost a decade ago "to warn against too slavish an acceptance of anybody's confidence intervals."[9] The word for the day is "humility."

REFERENCES

1. C. W. Bischoff, "Lags in Fiscal and Monetary Impacts on Investment in Producers Durable Equipment," Cowles Foundation Discussion Paper No. 250, June 11, 1968.
2. ———, "Hypothesis Testing and the Demand for Capital Goods" (draft of Nov. 20, 1968).
3. R. Eisner, "A Distributed Lag Investment Function," *Econometrica,* Jan., 1960, pp. 1–29.
4. ———, "A Permanent Income Theory for Investment: Some Empirical Explorations," *A. E. R.*, June, 1967, pp. 363–90.
5. ———, "Tax Policy and Investment Behavior: Comment," *A. E. R.*, June, 1969.
6. R. Eisner and M. I. Nadiri, "Investment Behavior and Neo-Classical Theory," *Rev. of Econ. Statis.,* Aug., 1968, pp. 368–82.
7. R. Eisner and R. H. Strotz, "Determinants of Business Investment," in Commission on Money and Credit, ***Impacts of Monetary Policy*** (Englewood Cliffs, N.J., 1963), pp. 59–337.
8. R. E. Hall and D. W. Jorgenson, "Tax Policy and Investment Behavior," *A. E. R.*, June, 1967, pp. 391–414.
9. D. W. Jorgenson, "Capital Theory and Investment Behavior," *A.E.R.,* May, 1963, pp. 247–59.
10. D. W. Jorgenson and J. A. Stephenson, "The Time Structure of Investment Behavior in United States Manufacturing, 1947–1960," *Rev. of Econ. Statis.,* Feb., 1967, pp. 16–27.
11. ———, "Investment Behavior in U.S. Manufacturing, 1947–1960," *Econometrica,* Apr., 1967, pp. 169–220.
12. J. R. Meyer and E. Kuh, ***The Investment Decision*** (Cambridge, 1957).
13. M. I. Nadiri and S. Rosen, "Interrelated Factor Demand, Adjustment Dynamics and Relative Prices," presented at Econometric Society Summer Meetings, Boulder, Colo., 1968 (forthcoming in the *A.E.R.*).
14. P. A. Samuelson, "Economic Forecasting and Science," ***Michigan Quar. Rev.,*** Oct., 1965, reprinted in P. A. Samuelson, J. R. Coleman, F. Skidmore, ***Readings in Economics*** (McGraw-Hill, 1967), pp. 124–29.

---

net stock of producers' durable equipment" has been introduced by Bischoff's adjustment of the relative price term, p/c. For his adjusted variable, here denoted as V, includes "a trend term which may reflect technical change, a systematic bias in the price index for capital goods due to failure to measure quality improvements, or a time trend in the relation between market variables used to measure the discount rate used by firms and the 'true' discount rate" [2, p. 9]. The wonder may be intensified by recognition that the $h = .00178$ plugged into the definition of $V_t = e^{ht}(p/c)_t$ is approximately the $h = .00182$ derived from nonlinear estimation of an equipment expenditure function reported by Bischoff in [1], where he notes (p. 69) that "*h* seems . . . to offset the trends in other variables."

[9] [7], p. 63.

INDUSTRIAL RESEARCH AND DEVELOPMENT: CHARACTERISTICS, COSTS, AND DIFFUSION OF RESULTS*

By EDWIN MANSFIELD
University of Pennsylvania

I. *Introduction*

In view of the size of the current R and D effort in the United States[1] and its effect on other economic variables like the rate of technological change, the rate of investment, and the rate of economic growth, it is important that we obtain a better understanding of the nature and characteristics of the R and D programs carried out by industry, the factors that seem to be associated with the cost of individual R and D projects, and the process by which the results of R and D become adopted and accepted. This paper summarizes very briefly some results of a number of related empirical studies of these topics.

II. *Characteristics of R and D Programs*

The first study, by Michael Hamburger and I,[2] is concerned with the characteristics of industrial research and development. Despite the attention that this area has received in recent years, little is known about the characteristics of the activities that firms call "research and development." This lack of information has been a hindrance to progress in this area, since, without a reasonable amount of information on this score, it is difficult to formulate or evaluate models relating R and D expenditures to other economic variables. In an attempt to shed additional light on this subject, we studied the characteristics of the R and D programs carried out in 1964 by twenty-two major firms in the chemical and petroleum industries—two industries that are very important performers of R and D.

The results indicate at least four things. First, they show that the bulk of the R and D projects carried out by these firms is relatively safe from a technical point of view. In practically all firms in our sample, most of the projects were regarded as having better than a fifty-fifty chance of technical success.[3] Second, only a small

---

* The studies described in this paper were supported by the National Science Foundation, the Ford Foundation, and the Small Business Administration. I want to express my appreciation to them for generous financial support.

[1] In 1966, industry alone performed about $15 billion of research and development. For a discussion of the growth of R and D expenditures in the postwar (and earlier) period, see Edwin Mansfield, *The Economics of Technological Change* (Norton, 1968), Chaps. 3 and 6.

[2] Edwin Mansfield and Michael Hamburger, "A Study of Research and Development in the Chemical and Petroleum Industries" (unpublished paper, 1968). Only a few of our results can be presented in the time available here.

[3] Of course, it might be objected that these estimates are biased upward because the R and D personnel are trying to sell projects. But we have data from related studies for some firms that allow us to compare estimated probabilities of success with actual probabilities of success. The results indicate that, although there may be some upward bias, it is not very large on the average. For example, in one firm, the average estimated probability of technical success was about 80 percent whereas the actual percentage of projects that were technically successful was about 75. In another firm, these figures were both about 80. Of course, projects that were not carried out as intended, or that were dropped for commercial reasons, were omitted to provide a fair comparison. A project is defined to be a technical success if it attains its technical objectives in the budgeted time and within the budgeted cost; otherwise it is defined to be a technical failure. Of course, such a simple classification can be misleading. For example a project may not attain its technical objective but it may nonetheless result in very valuable information. Despite these disadvantages, it seemed worthwhile to gather data concerning the estimated probability of technical success, since this estimate is commonly used by firms as a rough measure of the riskiness of a project from a technical angle, and it is likely to be as good a measure of this sort as one can obtain in a survey.

percentage of the money spent by these firms goes for basic research. On the average 9 percent of their R and D expenditures were devoted to basic research, 45 percent were devoted to applied research, and 46 percent were devoted to development.[4] Third, most of the R and D projects are expected to be quite profitable if they are technically successful, the median expected rate of return for a firm's projects generally being about 30 percent.[5] Fourth, the bulk of the R and D projects are expected to be finished and to have an effect on profits in five years or less.

These findings seem to support the hypothesis advanced by Hamberg[6] and others, that the bulk of the R and D carried out by large corporations is relatively safe and aimed at fairly modest advances in the state of the art. Models or policies based on the popular supposition that the bulk of the research and development in the industrial laboratory is very risky, far-out work aimed at really major inventions are likely to be misconceived and misleading. This is an important point, both for analysis and policy.

The results also indicate that, in each industry, there is a great deal of interfirm diversity in the characteristics of the R and D programs.[7] For example, some chemical firms devote 18 percent of their R and D expenditures to basic research, whereas others devote none. Some chemical firms spend $47,000 on R and D per scientist and engineer, whereas others spend $25,000. The median expected rate of return is about 40 percent for some chemical firms and less than 20 percent for others. The median length of time to completion is about two years for some chemical firms and about four years for others. Only in the case of the median estimated probability of technical success of projects is the interfirm variation in each industry relatively small. In all but three firms it lies between 50 and 62 percent in chemicals; in all firms, it lies between 70 and 83 percent in petroleum.[8]

III. *Size of Firm and Characteristics of R and D Programs*

Some of this variation among firms in the characteristics of their R and D programs can of course be explained by differences in size of firm.[9] In particular, one would suppose that the largest firms would devote a larger proportion of their R and D expenditures to more basic, more risky, and longer-term R and D projects than their smaller competitors. Our data for the chemical and petroleum industries bear out these hypotheses. For example, consider a chemical firm with sales that are one-tenth of the sales of the largest firm in the sample. Such a firm devotes to basic research a percentage of total R and D expenditures that is about one-third of the percentage devoted by the largest firm in the sample. Similarly, such a firm has a median probability of technical success that is about 20 percent higher, and a median length of time to completion that is about 15 percent lower, than the largest firm in the sample.

---

[4] We used the National Science Foundation's definitions of basic research, applied research, and development. For example, see *Basic Research, Applied Research, and Development in Industry, 1962* (NSF, 1965).

[5] Note that these estimates are conditional on the project's being technically successful. The unconditional estimates would, of course, be lower.

[6] Daniel Hamberg, "Invention in the Industrial Research Laboratory," *J.P.E.*, Apr., 1963.

[7] Of course, some of this variation could be due to differences among firms in the interpretation of the definitions of research and development, basic research, applied research, development, scientist, engineer, and so on. We questioned a number of the respondents to be sure that they were fully aware of the definitions and that they interpreted them properly. We could find no evidence that they were interpreting the definitions in significantly different ways.

[8] Note that we are concerned here only with the technical risks. Since technical success by no means insures commercial success, the total risks are considerably greater than indicated by the probabilities of technical success. But this does not alter the fact that the bulk of the projects tends to be relatively safe from a technical angle. For similar data for a large electrical equipment manufacturer, see Edwin Mansfield, *Industrial Research and Technological Innovation* (Norton, for the Cowles Foundation for Research in Economics, 1968), Chap. 3.

[9] For studies of the effect of firm size on total R and D expenditures, see Edwin Mansfield, *Industrial Research and Technological Innovation, op. cit.;* and Richard Nelson, Merton Peck and Edward Kalachek, *Technology, Economic Growth, and Public Policy* (Brookings, 1967).

However, it is very important to note that, although the largest chemical firm in the sample differs significantly from the relatively small firms in these respects, the difference between the largest firm and the average firm that is about one-half of its size, is generally small, if it exists at all. With regard to the percentage of total R and D expenditures devoted to basic research, the difference is only a couple of percentage points. With regard to the median probability of technical success and the median length of time to completion, the median probability is higher (not lower), and the median length of time is lower (not higher), in the largest firm in the sample than in a firm that is one-half its size. The results are much the same for the petroleum firms. Only in the case of the median length of time to completion is there a substantial difference between the largest petroleum firm in the sample and the average firm of about one-half its size.

Thus, the results indicate that firm size is associated with a substantial amount of the interfirm variation in these characteristics. They also indicate that, although the differences between the largest firms in the sample and the relatively small firms are sometimes quite large, the differences between the largest firms and firms of one-half their size are seldom large, if they exist at all. To the extent that the sample is trustworthy, the results suggest that, in these industries at least, firms as large as the largest firms in the sample are not required to insure that the existing amount of R and D of a more basic, risky, and long-term nature is carried out. Firms that are about one-half as large as the largest firms in the sample invest about the same percent of their R and D budget in more basic, risky, and long-term projects as do the largest firms in the sample. This too is an important point.[10]

IV. *The Costs of Development*

Let's turn from the characteristics of a firm's R and D program to the characteristics of an individual R and D project. An important, and extremely difficult, question is: what determines how much it costs to develop a new product? Some work bearing on this question is being carried out by Jerome Schnee,[11] who has obtained detailed data concerning about sixty development projects conducted by one major ethical drug firm between 1950 and 1967. Specifically, he has information concerning the costs of clinical testing and product formulation for each drug, and he is interested in determining whether it is possible to explain a reasonable amount of the variation among drugs in these costs by the characteristics of the product being developed, the development strategy, and so on.[12]

---

[10] It is important to note that we do not have complete data for all of the largest firms in these industries. Although there is no obvious bias in one direction or the other, the results pertain to only a sample of the firms, and it is always possible that this sample is unrepresentative in important ways. However, a reasonably large proportion of the firms are included, over 50 percent of the chemical firms with 1963 sales of $100 millions or more and over 50 percent of the petroleum firms with 1963 sales of $500 millions or more being included. Of course, nothing can be deduced from these results concerning other industries. Whether or not the same kind of results hold for other industries can only be determined by obtaining this sort of data for them.

[11] Schnee is working on a doctoral dissertation at the University of Pennsylvania. The results described in this section pertain to only part of his work and are tentative, since the project is still under way. The only other econometric studies of the determinants of development costs (of which I am aware) have been carried out at the RAND Corporation and pertain only to military development. Unfortunately, many of them are classified.

Three types of development projects are treated separately, since the nature of the development work is quite different in each type. These three categories are development projects that are aimed at (1) new chemical entities, (2) compounded products, and (3) alternate dosage forms. In each category, it is hypothesized that the costs depend on the number of dosage forms (tablets, oral liquid, nasal spray, etc.) being developed, which dosage forms are being developed, the therapeutic classification of the drug, and the spectrum of the drug's activity.[13] Also it is hypothesized that there has been an upward trend in costs during this period, due to inflation and changes (partly reflecting FDA regulation) in development procedures.[14] In addition, it is hypothesized that the costs will depend on the priority of the project and on whether or not parallel development efforts are employed. Finally, a variable similar to that used by Peck and Scherer[15] and Marschak, Glennan and Summers[16] to measure the extent of the advance in the state of the art was also included, more or less as an experiment.[17]

In each category, a regression based on these variables can explain a substantial portion of the observed variation in costs, R^2 being 0.93 for new chemical entities, 0.67 for compounded products and 0.90 for alternate dosage forms. In part, of course, these high correlations may be due to the fact that the data pertain to only one firm. Not all of the independent variables turn out to be statistically significant. Among new chemical entities, the number of dosage forms, the priority attached to the project, and calendar year are significant variables. Among compounded products, the spectrum of activity, the state of the art advance, calendar year, the therapeutic classification, and the use of parallel development efforts are significant variables. Among alternate dosage forms, the type and number of dosage forms, the therapeutic classification, the state of the art advance, and calendar year are significant variables.

---

[12] Although the costs of clinical testing of the compound and of product formulation are important parts of the total R and D costs pertaining to a drug, they are not all of the R and D costs associated with a particular drug. There are large costs that are usually incurred before the right compound is found. This should be borne in mind in interpreting the results.

[13] Dummy variables are used to represent which dosage forms are being developed and the therapeutic classification of the drug. The "spectrum of activity" refers to the range of a drug's biological activity and is measured by a dummy variable showing whether the drug is being developed for more than one therapeutic market. Some of these variables were omitted in particular categories. Among new chemical entities, the variable representing which dosage form was being developed was omitted because it seemed much less important than the number of dosage forms and to conserve degrees of freedom. Among alternate dosage forms, the spectrum of activity was omitted since they all were developed for a single market. One would expect that, other things equal, development cost would increase with the number of dosage forms and with the extent of the spectrum of activity. Also, some types of dosage forms are more expensive than others to develop.

[14] The calendar year is used as an independent variable. With regard to the FDA, 1962 amendments to the Food, Drug and Cosmetic Act strengthened the regulatory authority of the FDA over the clinical testing of new drugs and outlined minimum requirements for studies to demonstrate the efficacy and safety of a new drug.

[15] Merton Peck and F. M. Scherer, *The Weapons Acquisition Process* (Harvard, 1962).

[16] Thomas Marschak, Thomas Glennan, and Robert Summers, *Strategy for R and D* (Springer-Verlag, 1967).

[17] This variable represented a rough judgment of scientists and project managers familiar with these drugs. Essentially, the procedure used to obtain these judgments was like that used by Summers and Peck and Scherer. Obviously this variable is crude, which may help to explain why it is not always statistically significant. One would expect, of course, that projects attempting large advances in the state of the art would cost more than less ambitious projects. Dummy variables are used to indicate whether parallel development efforts were used and whether the project was given high priority by top management. Data concerning these variables could be obtained from the firm's records. On a priori grounds, it is difficult to predict the effect of the priority variable since costs tend to be high if there is a great emphasis on time reduction and also if there is very little emphasis on time reduction. The results, where significant, suggest that high priority projects are cheaper. The expected effects of the use of parallel development efforts are discussed below.

The results concerning the effects on costs of using parallel development efforts are quite interesting. Based on the theory underlying the use of parallel efforts, one would expect that parallel efforts would be more likely to reduce costs for new chemical entities than for compounded products or new dosage forms, since the uncertainty as to the optimal approach tends to be much greater in the former case. The results are consistent with this hypothesis, the estimated effect of parallel efforts being to increase costs among alternate dosage forms and compounded products and to decrease costs among new chemical entities. However, the effect of this variable is statistically significant only among compounded products.[18]

V. *Diffusion of the Results of Research and Development*

The results of industrial R and D, to be economically important, must be accepted and used. The final study I shall discuss is concerned with this diffusion process. Continuing along lines set forth in previous studies, I have carried out an investigation of the diffusion of numerical control in the tool and die industry, the purpose being to see how rapidly this innovation is spreading, the kinds of firms that are relatively quick to adopt it, the reasons given by firms for not using it, and the opinions of firm owners concerning the impact of the innovation on the structure of the industry.[19]

Numerical control—a way of operating machine tools by means of numerical instructions on tapes or cards—is certainly one of the most important innovations in this century. According to one leading research institute, numerical control "is the most significant new development in manufacturing technology since Henry Ford introduced the concept of the assembly line."[20] One of the industries most affected by this innovation is the tool and die industry. This industry is composed of a very large number of small firms, which means that we have the opportunity to study the diffusion of an innovation in an industry where there is very little concentration and where the organization of decision making in the firm is relatively simple. Other diffusion studies concerned with manufacturing have dealt chiefly with more concentrated industries.

To estimate the growth over time in the percent of tool and die firms using numerical control, a mail survey was carried out of the membership of the National Tool, Die, and Precision Machining Association.[21] The results indicate that less than 1 percent of the firms had begun using numerical control before 1961. By the beginning of 1966, the percentage had grown to 10 percent; and by the beginning of 1968, 20 percent of the firms were using numerical control. Allowing for differences in the profitability of using the innovation and the size of investment required, this innovation seems to be spreading more rapidly than innovations in other industries for which data are available. This is a noteworthy finding, because it supports the view that, all other things equal, innovations tend to spread more rapidly in less concentrated industries.[22]

---

[18] For discussion of the effect of parallel development efforts, see Burton Klein, "The Decision Making Problem in Development," *The Rate and Direction of Inventive Activity* (N.B.E.R., 1962); Richard Nelson, "Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts," *Rev. of Econ. and Statis.*, 1961; and Marschak's work in Marschak, Glennan, and Summers, *op. cit.*

[19] Only a few of the findings can be presented in the available time. For a more complete discussion, see Edwin Mansfield, "Numerical Control in the Tool and Die Industry: The Diffusion of a Major Technological Innovation" (unpublished paper, 1968). For a review of the literature, see E. Mansfield, *The Economics of Technological Change, op. cit.*, Chap. 4.

[20] Illinois Institute of Technology Research Institute, *Technological Change: Its Impact on Metropolitan Chicago* (1964), p. 1.

[21] Note that all of the findings pertain only to the approximately 1,000 members of the Association, since the costs of extending the frame beyond the Association membership were out of the question. The Association members account for the bulk of the industry's output. Figures for all members of the Association for 1967 and 1968 were available from the Association's *Directory*, so these figures contain no sampling error. A carefully selected sample of firms was interviewed. The results, which are unbiased, are in close agreement with the results of the mail survey.

VI. *Characteristics of Early Users*

There are a good many reasons for expecting the larger tool and die firms to be quicker than the smaller ones to introduce numerical control. For example, the larger firms are more likely to have the financial resources to enable them to experiment, and they are more likely to have the technical know-how and the managerial qualities that are important in determining a firm's speed of response to a new technique. This hypothesis is borne out by my results. The median employment of firms using numerical control at the beginning of 1968 was about seventy, while the median employment of nonusers was about twenty-five. Moreover, among the users of numerical control, there is a significant inverse relationship between the size of a firm and the year when it began using numerical control.[23]

Other variables that would be expected to influence whether or not a firm adopted numerical control before 1968 is the education and age of the firm's president. Better educated entrepreneurs are likely to be in a better position to understand the issues regarding numerical control, to have the flexibility of mind to use it, and to be in contact with technical and university centers and the relevant literature. Younger entrepreneurs would be more likely to make the break with the past, their emotional attachments to old skills and old technology being weaker and their willingness to take risks probably being greater than their older rivals. The data are consistent with these hypotheses. Most of the users (for which we have data) are college graduates while most of the nonusers finished high school or less; the median age of the users was about forty-eight while the median age of the nonusers was about fifty-five. However, when a multiple regression is run (age and education being independent variables, the dependent variable being a dummy variable showing whether or not a firm used numerical control before 1968), the effect of education is statistically significant, but the effect of age is not.

A carefully selected sample of twenty-eight firms without numerical control was interviewed to determine why they were not using it. Practically all claimed that it would not be profitable for them at present, the primary reason being that they do one-of-a-kind work. (About 20 percent of the firms planned to begin using it in the next year or two.) However, about 30 percent of the managers, by their own account, had important gaps in their knowledge and understanding of numerical control. About 10 percent of the firm owners, being close to retirement, had decided to stick with conventional methods until they retired—or until they were put out of business. Judging by the interviews and other evidence, the diffusion process seems to have been slowed perceptibly by misunderstanding of the innovation and resistance to change.[24]

---

[22] This finding is in agreement with my previous results, which pertain to four other industries. They indicate that, holding constant the profitability of the innovation and the size of the investment, the rate of imitation tends to be greater in less concentrated industries. However, there were too few industries to be at all sure the relationship was not due to chance. See E. Mansfield, ***Industrial Research and Technological Innovation, op. cit.***, Chap. 7.

[23] To some extent, this relationship is probably due to a difference between the growth rates of users and nonusers of numerical control, users probably having had higher growth rates. But this can explain only part of the relationship, there being evidence that, when they began using numerical control, the early users tended to be larger than the nonusers. Of course, the fact that the larger firms tend to be quicker than the small firms to begin using an innovation does not contradict in any way the finding that the rate of imitation tends to be faster in less concentrated industries.

VII. *Conclusion*

In conclusion, organized scientific and inventive activity is a relatively new, and very important, factor in the modern economy. This paper has presented a brief summary of the results of several empirical studies of industrial research and development, these studies focusing on its characteristics, cost, and diffusion. Some of the principal findings are the following: First, the bulk of the R and D projects carried out by a sample of major chemical and petroleum firms seems relatively safe from a technical point of view, most projects having better than a fifty-fifty chance of technical success. This supports the hypothesis advanced by some economists that the bulk of the R and D carried out by large corporations is relatively safe and aimed at fairly modest advances in the state of the art.

Second, the largest firms in the sample devote a larger proportion of their R and D expenditures to more basic, more risky, and longer-term projects than their smaller competitors. However, although the differences between the largest firms in the sample and relatively small firms are sometimes quite large, the difference between the largest firms and firms of one-half their size are seldom large, if they exist at all. To the extent that the sample is trustworthy, the results suggest that, in these industries at least, firms as large as the largest firms in the sample are not required to insure that the existing amount of R and D of a more basic, risky, and long-term nature is carried out. Firms that are one-half as large as the largest firms in the sample invest about the same percent of their R and D budget in more basic, risky, and long-term projects as do the largest firms in the sample.

Third, a detailed study of about sixty development projects carried out in a large ethical drug firm indicates that the cost of developing a particular product is related to the therapeutic classification of the product, its spectrum of activity, and the number and type of dosage forms. It is also related to the priority attached to the development project and to whether or not parallel development efforts are used, at least in some product categories. This seems to be the only econometric study of the determinants of development costs in the civilian economy that has been carried out to date. Much more work is needed.

Fourth, a study of the diffusion of numerical control in the tool and die industry indicates that, allowing for differences in profitability and size of investment, this innovation seems to be spreading more rapidly than innovations in other industries for which data are available. This finding tends to support the hypothesis that innovations generally spread more rapidly in less concentrated industries. The early users of numerical control tend to be the larger tool and die firms. Also, whether or not a firm is using numerical control seems to be related to the age and education of the firm's owner; but when both age and education are introduced simultaneously, only the effect of education is significant.

---

[24] Nonetheless, as noted in Section V, the imitation process seems to be going on relatively rapidly. Thus, these drags on the rate of imitation may be less important than in other cases for which we have data.

MARKET STRUCTURE AND THE STABILITY OF INVESTMENT*

By Frederic M. Scherer
University of Michigan

This paper reports on an empirical venture into the darkest *terra incognita* of industrial organization analysis: the links between market structure and dynamic stability. Concretely, it seeks to determine whether capital investment expenditures were more stable between 1954 and 1963 in concentrated or in atomistically structured manufacturing industries, other things being equal.

One can find numerous a priori conjectures on this subject, some suggesting that concentration is relatively conducive to stability, others the opposite. Representative of the first group is Tibor Scitovsky's assertion that pure competitors overreact to demand shifts because they ignore the direct and indirect effects of their investment on price. Oligopolists and monopolistic competitors are said to be less prone to overadjustment because they suffer no illusions about facing an unlimited market.[1] In a similar vein is G. B. Richardson's sophisticated new variant on the conventional cartel wisdom, holding that "contrived" market imperfections—overt price-fixing arrangements, large-scale mergers, and conscious parallelism—create an information system which helps call forth the proper amounts of investment by the right firms in response to demand changes.[2]

Others have implied that investment is likely to be particularly unstable under oligopolistic conditions. According to James Duesenberry, market position among oligopolists refraining from price rivalry depends upon such nonprice variables as reliability of service and personal buyer-seller ties. The firm with reserve capacity to serve customers in times of peak demand stands a good chance of retaining their loyalty permanently, and this motivates each oligopolist to try increasing its market share through competition in production capacity.[3] To carry the logic a step beyond Duesenberry's explicit statement, fear of losing market share and future profit-earning potential to more aggressive rivals can lead oligopolistic producers to invest heavily in additional capacity when demand is growing. They may then cut back sharply when serious excess capacity appears. The net result is a volatile investment profile over time. In Joe S. Bain's somewhat different view, investment will tend to be unstable when barriers to new entry are modest while industry structure is sufficiently concentrated to support prices temporarily above the competitive level. Then producers may strive to maximize short-run collective profits, encouraging a wave of new investment which ends abruptly when the price structure collapses.[4] A third possibility is that concentrated industries, with generally higher profits and cash flow, are better able to finance investment programs when funds markets are tight, and therefore will respond to increases in demand with a sharper acceleration of in-

---

* Work on this paper was supported by the Institute of Public Administration, University of Michigan. The author is indebted to Lowell M. Seyburn for research assistance, and to W. H. Locke Anderson for critical comments.

[1] Tibor Scitovsky, *Welfare and Competition* (Irwin, 1951), pp. 365–66.

[2] G. B. Richardson, *Information and Investment* (Oxford: Oxford Univ. Press, 1960), pp. 67–70 and 128–36.

[3] James S. Duesenberry, *Business Cycles and Economic Growth* (McGraw-Hill, 1958), pp. 113–33. See also Donald H. Wallace, *Market Control in the Aluminum Industry* (Harvard Univ. Press, 1937), pp. 336–43.

[4] Joe S. Bain, *Barriers to New Competition* (Harvard Univ. Press, 1956), pp. 33–41 and 189–90.

vestment. Finally, investment may be more unstable in concentrated industries simply because there are fewer independent centers of entrepreneurial initiative, and hence less opportunity for the spending peaks of one firm to be offset by troughs in other firms' outlays.

The Basic Model

Traditional supply and demand influences undoubtedly dominate investment decisions, and market structure is likely to have an impact only at the margin. To detect the market structure effect, a large cross-section of well-defined industries appears necessary. Given current data limitations, this requirement is incompatible with the estimation of an investment model firmly rooted in economic theory, so my main thrust at resolving the clash of a priori conjectures takes a less elegant, more roundabout tack.

From the biennial Census and annual surveys of manufactures, estimates of new capital expenditures for numerous four-digit manufacturing industries are available, at least from 1954 on. The index of investment instability V_i to be used as principal dependent variable in the subsequent analysis is the standard error of estimate obtained after regressing the logarithms of annual new capital expenditures between 1954 and 1963 for each sampled industry on a simple time trend variable. Because V_i is scaled logarithmically, it reflects proportional deviations from trend investment, and therefore affords a directly comparable index of instability between industries of varying size. The higher V_i is, the more unstable observed investment was relative to a smooth time trend.

The main sample covers eighty four-digit manufacturing industries originating 37 percent of all new capital investment in manufacturing during 1963. The sample is exhaustive, subject to the satisfaction of two criteria. First, the investment series for the ten years covered had to be complete or nearly complete; and it had to be continuous or marred by a spliceable discontinuity (i.e., due to minor industry definition changes) not exceeding 4 percent of trend outlays. Industries were included despite a lack of investment data for one year in twenty-one cases, and in one case an industry without data for two years was accepted. The remaining industries had complete data series. Second, the Census industry definition had to be such that four-firm concentration ratios could be used either directly or after feasible adjustments as a meaningful index of structural market power. Many otherwise complete data sets were thrown out under this criterion. For nine included industries, concentration ratios were adjusted upward because component five-digit product lines had greater economic relevance or because the markets were predominantly regional. Eight industries defined too narrowly were consolidated into three more meaningful groups. Through this latter adjustment, the number of separately defined industry data sets was reduced to seventy-five.

The independent variable of paramount concern here is market concentration C_i, measured by the simple average of the unadjusted or (in twelve industries) adjusted four-firm concentration ratios for 1954, 1958, and 1963. Several additional independent variables were introduced to take into account other factors affecting the variability of investment.

Investment outlays are likely to be more unstable, the more unstable demand is over time. One approach toward reflecting the variability of demand is to employ the variable VM_i, which is the standard error of estimate derived after regressing the logarithms of annual production worker man-hours between 1954 and 1963 on a time trend. Man-hours were chosen over the sales variable commonly used in accelerator analyses because the variability of sales over time is affected more strongly by differences in pricing behavior associated with market structure. An alternative or complementary approach which, we shall see, performs better statistically, is to introduce dichotomous dummy variables reflecting relevant product characteristics. Three such variables were used: Dur_i distinguishes durable from nondurable goods; $Cons_i$ distinguishes consumer goods from other commodities; and Int_i distinguishes intermediate goods (such as steel, flour, and cement) from commodities sold as end items. Since the demand for durables is notoriously unstable, we predict a strong positive relationship between V_i and Dur_i. The demand for consumer goods is probably more stable than for other commodities, suggesting a negative association between V_i and $Cons_i$; while the effect of Int_i is not obvious on a priori grounds.

Observation suggests that capital in capital-intensive industries is lumpier and more clay-like, implying greater instability of outlays over time. Also, if plant and equipment in the more capital-intensive industries are more durable, and if replacement investment has not attained its steady-state equilibrium growth path, accelerator-induced fluctuations in expansion investment will be larger relative to total investment (expansion plus replacement) in the more capital-intensive fields. To reflect these possibilities, we introduce a variable $(K/VA)_i$, defined as the ratio of December, 1957, capital investment at book value to 1958 value added.

Finally, we shall employ an industry size variable S_i, defined as the simple average of the logarithms of value added in 1954, 1958, and 1963. We anticipate a negative relationship between V_i and S_i because large industries tend to include more independent producers, to have more diversified product lines, and to face more steeply rising short-run capital goods supply functions when they attempt to expand investment rapidly.

Summing up, the principal variables employed are as follows:

V_i	Variability of 1954–63 investment from its time trend.
C_i	The four-firm concentration ratio.
VM_i	Variability of 1954–63 man-hours from their time trend.
Dur_i	A durable goods dummy variable.
Int_i	An intermediate products dummy variable.
$Cons_i$	A consumer goods dummy variable.
$(K/VA)_i$	The ratio of 1957 capital assets to 1958 value added.
S_i	Industry size: mean of the 1954–63 value-added logarithms.

The Principal Results

Table 1 presents a correlation matrix covering the most important variables. The positive correlation of 0.331 between

TABLE 1

CORRELATION MATRIX OF THE PRINCIPAL VARIABLES; $N=75$*

	V_i	C_i	Dur_i	S_i	$(K/VA)_i$	VM_i
V_i	1.000	.331	.354	−.299	.288	.309
C_i		1.000	.108	−.209	.116	.205
Dur_i			1.000	.082	.066	.510
S_i				1.000	.256	−.083
$(K/VA)_i$					1.000	−.076
VM_i						1.000

* The 5 percent point is .225; the 1 percent point is .292.

V_i and concentration is significant at the 1 percent level. All signs of the simple correlations between V_i and the other independent variables conform to a priori expectations. It is noteworthy that with only one exception (involving Dur_i and VM_i) the independent variables are more highly correlated with V_i than with each other.

The linear regression of V_i on all the main independent variables except VM_i is as follows, where concentration C_i is scaled in ratio form:

$$(1)\qquad \begin{array}{llllll} V_i = .181 + & .057\,C_i + & .054\,Dur_i + & .0035\,Cons_i + & .022\,Int_i & \\ & (.029) & (.018) & (.0186) & (.017) & \\ & [.232] & [.345] & [.023] & [.157] & \\ - & .055\,S_i + & .056\,(K/VA)_i; & & R^2 = .407; & N = 75. \\ & (.015) & (.023) & & & \\ & [-.404] & [.280] & & & \end{array}$$

(Standard errors are given in parentheses under the coefficients, and partial correlation coefficients are bracketed.) The concentration coefficient is significantly different from zero at the 6 percent level in a two-tail test. The signs of the other major independent variables, with the exception of $Cons_i$, support the a priori predictions.[5]

When the standard error of trend-adjusted production worker man-hours VM_i is introduced as an additional index of demand instability, the results are as follows:

$$(2)\qquad \begin{array}{lllllll} V_i = .160 + & .050\,C_i + & .045\,Dur_i + & .0099\,Cons_i + & .027\,Int_i & & \\ & (.029) & (.019) & (.0190) & (.017) & & \\ & [.204] & [.284] & [.064] & [.188] & & \\ - & .054\,S_i + & .059\,(K/VA)_i + & .625\,VM_i; & & R^2 = .424; & N = 75. \\ & (.015) & (.023) & (.437) & & & \\ & [-.402] & [.297] & [.172] & & & \end{array}$$

In this regression, the concentration coefficient is significant only at the 10 percent level in a two-tail test, while the coefficient for VM_i is significant at the 8 percent level in a one-tail test. However, since the regression coefficients are adversely affected by modest multicollinearity among C_i, Dur_i, VM_i, and S_i, I am inclined to interpret the regression results as fairly strong support for the hypothesis that investment is more variable in concentrated than in atomistic industries. Raising the four-firm concentration ratio from .10 to .90—i.e., moving from an atomistic structure to tight oligopoly—leads to a 40 percent increase in the instability of investment as estimated by equation (1) when all other variables are held at their mean values, or to a 34 percent increase according to equation (2).

These results persist under a variety of sensitivity tests. When the eight industries with the least satisfactory market definitions were deleted from the sample, the partial correlation coefficient for concentration in the analogue of equation (1) increased to 0.297 and the t-ratio of its regression coefficient rose to 2.41. Replacing the concentration index with a variable counting the number of firms in each industry led to a pronounced drop in R^2, even though the two were correlated, with $r=-0.397$. In a regression incorporating both variables, concentration had by far the more potent explanatory effect. Since it is easier to draw an exhaustive sample in a concentrated industry, the regression estimates might conceivably have been affected by pure sampling error. However, introduction of sampling error estimates from the 1959–60 Survey of Manufactures as an additional variable had no perceptible effect on the regressions, though the sampling errors were negatively correlated with concentration, with $r=-0.17$.

---

[5] The simple correlation between $Cons_i$ and V_i is negative, as predicted.

Under my interpretation of Duesenberry's conjecture, the instability of investment should be greatest in oligopolistic industries. The introduction of a squared concentration term to test for nonlinearity yielded a minute and statistically insignificant incremental increase in R^2. Failure to reject the linearity hypothesis does not necessarily show that investment is more unstable under pure monopoly than under oligopoly, since the sample included only one industry (telephone equipment) virtually dominated by a single firm.[6] The results do, however, suggest that investment is more unstable in tight than in loose oligopolies.

Cyclical or Random Instability?

We find then that investment tends to deviate more from its trend values when concentration is high than when it is low. It remains unsettled whether these deviations are of a cyclical or largely random character. Cyclical instability would prevail if oligopolists increased their capital outlays especially rapidly in response to demand increases, cutting back with equal vigor during slumps. Alternatively, their outlays might be scattered widely but in essentially random fashion across the full business cycle, without notably greater dispersion at peaks and troughs.

To shed some light on these alternatives, Durbin-Watson autocorrelation coefficients were computed from the time trend regression equations of the fifty-three industries on which complete investment data were available. For this subsample, the Durbin-Watson statistics were negatively correlated with V_i, the simple correlation coefficient being -0.432.[7] They were negatively but insignificantly correlated with concentration, with $r=-0.131$. The effect of concentration in multiple regressions with the Durbin-Watson coefficients as the dependent variable continued to be negative but statistically insignificant, with t-ratios well below 1.0. This suggests that investment was at best only slightly more volatile cyclically in the more concentrated industries, and that the observed instability of concentrated industry investment was mainly of a more or less random character. Unusually strong cyclical bunching of capital outlays for the concentrated markets was not evident.

Analysis of interactions between the concentration and industry size variables supports this interpretation. Consistent with Ralph L. Nelson's studies of concentration ratio biases,[8] C_i and S_i are negatively collinear, with $r=-0.209$. They compete for explanatory power. When, for example, S_i is introduced into a regression of V_i on the three dummy variables and C_i, the concentration variable's regression coefficient t-ratio falls from 3.09 to 2.53. This suggests that they must in part be explaining the same thing: the differences in investment variability stemming from some combination of size, diversity, and diffusion of decision-making foci. If in fact the joint effect of low concentration and high industry size primarily reflected success in hedging against otherwise uncontrollable demand fluctuations, we should expect it to operate analogously in explaining the intertemporal variability of production worker man-hours. But a regression of VM_i on the relevant variables was as follows:

---

[6] It is worth noting that the observed value of V_i for telephone equipment exceeds the value predicted from equation (1) by 28 percent.

[7] There is a spurious element in this correlation, since the denominator of the Durbin-Watson statistic is identical to the numerator of the standard error of estimate formula. It is worth noting also that the results observed in equations (1) and (2) persist for the fifty-three industry subsample, although the explanatory power of concentration is stronger in the subsample, with a regression coefficient t-ratio of 2.58 for the analogue of equation (1).

[8] Ralph L. Nelson, *Concentration in the Manufacturing Industries of the United States* (Yale Univ. Press, 1963), pp. 41–48.

$$(3)\; VM_i = .035 + \underset{\substack{(.005)\\ [.328]}}{.014}\, Dur_i - \underset{\substack{(.0051)\\ [-.226]}}{.0097}\, Cons_i - \underset{\substack{(.0042)\\ [-.236]}}{.0085}\, Int_i + \underset{\substack{(.0077)\\ [.153]}}{.0099}\, C_i - \underset{\substack{(.0039)\\ [-.079]}}{.0026}\, S_i;$$

$$R^2 = .344; \quad N = 75.$$

Industry size does compete with concentration, reducing the latter's t-ratio from 1.47 without S_i to 1.28 with it, but the size variable no longer overshadows concentration, as it did in equations (1) and (2). It would appear therefore that the joint effect of C_i and S_i operates mainly with respect to investment and not with respect to production worker employment. Decisions affecting the flow of investment are no doubt subject to greater managerial discretion than decisions concerning the size of the work force, which more nearly resembles a stock variable. And when concentration is low and industry size large, there is a stronger relative tendency for random deviations in the exercise of that discretion to cancel one another out.

One further loose end must be pinned down. It is conceivable that concentrated industries have a different appropriation-to-expenditures lag pattern than atomistically structured industries, and this could affect the results. Cyclical bunching would be especially likely when the bulk of investment outlays occur within a short (but lagged) time period, once the decision to invest has been taken in response to a market stimulus. To explore this possibility a "coefficient of response sharpness" was computed from Mrs. Almon's study of the appropriation-to-expenditures distributed lag pattern in fourteen two-digit manufacturing industry groups.[9] It measures the proportion of expenditures occurring in the four contiguous quarters of most intensive lagged expenditure. Two indices of structural market power were also compiled: a value-added weighted average index of four-firm sales concentration in the component four-digit industries for 1954 (hereafter, the "unadjusted concentration index"); and the analogous index developed by Leonard Weiss after adjustments were made in the raw four-digit industry data to correct for market definition inadequacies (hereafter, the "Weiss concentration index").[10]

The simple correlation between the coefficient of response sharpness and the unadjusted concentration index was -0.414. Using the Weiss concentration index, the correlation was -0.316. Neither correlation is significantly different from zero at the 10 percent level. This result suggests that the appropriation-to-expenditure response pattern is not sharper or more peaked in concentrated than in atomistic industries, and it may well be less sharp. It implies in turn the absence of technological reasons for greater cyclical bunching of concentrated industry investment outlays in response to a demand stimulus. If concentrated industries exhibit greater cyclical volatility of investment, it must be because they are quicker on the appropriations trigger.

---

[9] Shirley Almon, "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica*, Jan., 1965, pp. 188–89.

[10] Leonard W. Weiss, "Average Concentration Ratios and Industrial Performance," *J. of Ind. Econ.*, July, 1963, p. 242.

A Further Test

The analysis thus far is based upon a naïvely specified model. A further test embodying a more sophisticated econometric structure was attempted, using the results obtained by Jorgenson and Stephenson in fitting their "neoclassical" distributed lag model to thirteen two-digit industry quarterly investment time series for the period 1947–60.[11] Two questions are of interest: whether the prediction errors generated by the Jorgenson-Stephenson model are larger in concentrated than in atomistic industries, and whether the estimated lag structure is more peaked or sharper in concentrated industries.

The measure of prediction error size chosen to test for the first relationship was the coefficient of residual variation from their regression equations; that is, the standard error of estimate divided by mean investment. The simple correlation between the coefficients of residual variation for the thirteen industries and the appropriate unadjusted concentration indices was +0.675. When the Weiss concentration indices were used, the correlation was +0.568. These correlations, which are significantly different from zero at the 5 percent level, reinforce my previous conclusion that the unsystematic variability of investment is greater in concentrated than in unconcentrated industries. The strength of the correlations is surprising, since if high variability in concentrated industries was mainly due to a dearth of independent decision-making foci, one would expect considerable attenuation when four-digit industries are aggregated into two-digit groups.[12]

To test for the second relationship, a coefficient of response sharpness, measuring the ratio of spending in the five quarters of most intensive lagged investment to total expansion investment, was computed. The simple correlation between it and the unadjusted concentration indices was −0.191, and between it and the Weiss concentration indices −0.166. Neither correlation is significant at the 10 percent level. The negative sign might be the result of technological influences, mirroring the negative association between the sharpness of appropriation-to-spending lags and concentration observed earlier. Indeed, the possibility of a sharper demand stimulus-to-appropriations lag in concentrated industries is not altogether inconsistent with the evidence, although this conjecture has not been tested directly.

Conclusion

Summing up, we find that investment outlays tend to be more unstable relative to their trend values in concentrated than in atomistically structured industries, *ceteris paribus*. This greater instability appears to be more the result of low decision-making power dispersion than excessive sensitivity to demand stimuli, although the latter effect cannot be ruled out completely. The results clearly provide no support to those who argue for concentration and cartelization because of alleged macroeconomic stabilizing properties. But the greater instability of concentrated industry investment may also not be very harmful socially, at least to the extent that it is largely of a random character, and not synchronized with aggregate business fluctuations. For although statistical averaging-out may not operate fully within specific concentrated industries and two-digit fields, there may be further variance attenuation at the macroeconomic level as noncyclical deviations in individual industry outlays are pooled across industry lines.

---

[11] Dale W. Jorgenson and James A. Stephenson, "The Time Structure of Investment Behavior in United States Manufacturing, 1947–1960," *Rev. of Econ. and Statis.*, Feb., 1967, p. 21; and *idem*, "Investment Behavior in U. S. Manufacturing, 1947–1960," *Econometrica*, Apr., 1967, p. 190. Mean industry investment data were supplied by Professor Jorgenson.

[12] An alternative explanation might be that the Jorgenson-Stephenson model, which assumes competitive product markets, is simply misspecified when concentration is high. Intertemporal changes in the elasticity of product demand affect the monopolist's investment marginal revenue product in a wider range of cases than they affect the competitive firm's VMP. However, in a world of oligopolies it seems more likely that the observed behavioral differences are attributable primarily to the structural links explicitly hypothesized in the text.

RESEARCH AND DEVELOPMENT, PRODUCTION FUNCTIONS, AND RATES OF RETURN

By JORA R. MINASIAN
University of Southern California

Almost a decade ago I argued that research and development activity represented efforts made towards achieving efficiency, and showed that the rate of growth in productivity of the firms was positively correlated with their research and development expenditures.[1]

This paper reports the results of a study of "rate of return" to research and development expenditures in a production function setting.[2] Although the tenuous nature of the Cobb-Douglas type of production function was originally indicated by two independent studies in the early 1960's[3] because of its predominance in use, the estimates are obtained by both a covariance model and by the use of a production function of the Cobb-Douglas type.

The Model

The production function of a firm may be expressed by equation (1).

$$V_{ft} = f(L_{ft}, K_{ft}, T_{ft}) \tag{1}$$

where V_{ft}, L_{ft}, K_{ft}, and T_{ft} represent, respectively, value added, labor, capital, and technology of the firm in time, t. The firm's technology in time, t, is a positive function of the existing stock of knowledge. We postulate that the research and development expenditures are related in additions made to the stock of knowledge. That is,

$$T_t = g(N_t) \tag{2}$$

and

$$N_t = h\left(\sum_{t=0}^{t=t} R \right) \tag{3}$$

where N and R stand for stock of knowledge and research and development expenditures. Substituting (2) and (3) in equation (1) we obtain the production function of the firm expressing the value added as function of labor, capital, and accumulated research and development expenditures of the firm in time, t. That is,

$$V_{ft} = k\left(L_{ft}, K_{ft}, \sum_{t=0}^{t=t} R_{ft} \right) \tag{4}$$

The Statistical Method of Analysis

Observations on the value added, capital, and labor during 1948–57 for seventeen firms in the chemical industry constitute the sample. We first visualize a production function relating value added to labor and capital but allowing for differences in the firms' technologies. This may be done by constant terms for each firm in the context of covariance analysis. These firm constants summarize the differences among firms. Therefore, we can pool cross-section and time series data of the firms to estimate the constants and the regression coefficients which are freed from the differences in the firms.

A postulated reason for such differences is the firms' research and development expenditures. Therefore, one implication is that if these firm constants were statistically significant when the production function omitted the research and development, they should tend to lose their significance in the presence of the research and development in the function.

---

[1] Jora R. Minasian, "The Economics of Research and Development," *The Rate and Direction of Inventive Activity: Social and Economic Factors*, A Conference of the Universities-National Bureau Research (Princeton Univ. Press, 1962), pp. 93–143.

[2] For a summary of partial results of a study using the covariance analysis, see my "Technical Change and Production Function," *RAND P-2430*, pp. 1–14, presented at the meeting of the Econometric Society, 1961, and printed in the Proceedings of *Econometrica*, Apr., 1962.

[3] Jora R. Minasian, "Elasticity of Substitution and Constant-Output Demand Curves for Labor," *J. P. E.*, June, 1961. Also K. Arrow, H. B. Chenery, B. S. Minas, and R. M. Solow, "Capital-Labor Substitution and Economic Efficiency," *Rev. of Econ. and Statis.*, 1961.

In general, firms may differ for reasons not incorporated in research and development variables. It is preferable for this reason to use a function which allows a constant term for each firm. Further, it is not necessary that the firm constants become insignificant as there is no a priori reason to believe that all potential differences among firms are perfectly correlated with research and development expenditures.

It should, therefore, be noted that the primary purpose of adding variables to the function is not to achieve higher multiple coefficients of determination; rather it is to distribute the pie, as it were, more appropriately and thus obtain a more accurate set of partial regression coefficients.

Thus, the methodology will be to employ a multiple regression technique to test an alternative hypothesis involving the inclusion of the variables in the regression equation and to shed light on the nature of the explanation provided by the variables.

Both capital and research and development are gross stock variables. The alternative of assuming that there is a constant relationship between service flows and the stock of capital and research and development was preferred to an attempt to correct the service flows by some arbitrary percentage representing presumably obsolescence, breakdowns, uselessness of knowledge, etc. It was felt that the errors involved in the former would be less than those present in the latter procedure.

Finally, note that the model specifies research and development undertaken by the firm alone. The benefits derived from knowledge generated by the industry and the economy are not explicitly incorporated in the model. The effect of the excluded variable will be captured by the constant term in the regression equation, if the excluded variable is not correlated with the firm's research and development. The returns to research and development will be overestimated in the case of a positive correlation between the excluded variable and the firm's research and development expenditures.

The Empirical Results

For each set of variables two types of production functions (linear in the logs) were estimated. In one there is a general constant and seventeen specific constants—one for each firm, and alternatively ten time constants, one for each year. We shall call this Type I formulation. In the other there is only one general constant term. We shall call this Type II formulation. Of course, both use combined cross-section and time series data.

The regression results of Type I formulation are given in Table 1, and those for Type II formulation appear in Table 2. To conserve space, firm and time constants are not given in Table 1. Instead a dagger is used to show which of the constants were estimated in a particular regression.

First, evaluation of time and firm constants. The independent variables explain the variation in the value added not explained by the constants. For example, adding K and L to the equation, that contains only the constants, yields an F-value of 350 which is highly significant; 1 percent level of significance requires an F-value of 2.20. Time constants are significant in regressions 2 and 3 but not in regressions 4 and 5. The F-values are, respectively, 4.0, 4.6, 1.12, and 1.72.

Firm constants are highly significant in all the regressions: 6, 7, 8, and 9; the F-values are 27, 17, 36, and 24, respectively.

TABLE 1

TYPE I FORMULATION

Multiple Regressions of the Value Added (V) on Labor (L), Capital (K), Research and Development (R), Time (T), Chemical Firms 1948–57

$$\log V_{ft} = \log b_0 + \log B_f + \log B_t + b_1 \log L_{ft} + b_2 \log K_{ft} + b_3 \log R_{ft} + b_4 \log T_{ft}$$

V_{ft}	b_0	B_f	B_t	L	K	R	T	R^2
1	− .119			.582 (.042) .734	.459 (.039) .678			.970
2	− .069		†	.621 (.040) .771	.418 (.038) .661			.974
3	.282		†	.382 (.042) .583	.357 (.032) .669	.258 (.029) .580		.983
4	.553		†	.621 (.040) .744	.418 (.037) .661		−.950* (2.988) .033	.975
5	.185		†	.382 (.043) .583	.357 (.032) .669	.259 (.029) .580	.150* (2.485) .001	.983
6	−1.173	†		.596 (.083) .505	.645 (.059) .667			.992
7	− .019	†		.820 (.076) .659	.156 (.081) .156	.113 (.015) .531		.995
8	.037	†		.825 (.077) .658	.212 (.076) .221		.124 (.016) .525	.994
9	.007	†		.823 (.076) .659	.167 (.084) .160	.084* (.068) .099	.033* (.075) .036	.995

See Appendix for definition of the variables.

In each regression, the partial regression coefficient is given in the first row, its standard error (in parentheses) in the second row, and the partial coefficients of determination in the third row. Partial regression coefficients that are not significantly different from zero at the 5 percent two-tail level are marked by *. Finally, † indicates the type of constant estimated but is not given in the table for economy of space.

In all of the cases, the unexplained variations in value added are reduced by a greater amount by firm than by time constants. The combined time and firm constants explain most when the regression equation includes T, but least when R is included in the regression equation. Finally, although statistically significant, the firm constants lose their explanatory power more with R than with other variables included in the regression equation. Adding R to the regression equation reduces the explanatory power of firm constants by 38 percent. In summary, then,

TABLE 2

TYPE II FORMULATION

Multiple Regressions of the Value Added (*V*) on Labor (*L*), Capital (*K*), Research and Development (*R*), Time (*T*), Chemical Firms 1948–57

$$\log V_{ft} = \log a_0 + a_1 \log L_{ft} + a_2 \log K_{ft} + a_3 \log R_{ft} + a_4 \log T_{ft}$$

V_{ft}	a_0	L_{ft}	K_{ft}	R_{ft}	T_{ft}	R^2
1	−.119	.582 (.042) .734	.459 (.039) .678			.970
2	.129	.522 (.036) .748	.368 (.034) .638	.127 (.015) .540		.990
3	−.130	.621 (.041) .766	.418 (.038) .651		.094 (.021) .326	.973
4	.386	.393 (.043) .577	.356 (.033) .649	.249 (.029) .550	−.172 (.036) .347	.982

See notes at the bottom of Table 1.

time constants are substitutes for *T*; compare the regression coefficients for *L* and *K* in regressions 2 and 3 with those in regressions 4 and 5, respectively, Table 1. Both are inferior to firm constant and *R* in explaining variation in value added.

Second, the results of the regression coefficients. Once firm constants are introduced, the regression coefficients of *L* and *K* are stable, whether *R* or *T* or both are included in the regression equation; see regressions 7, 8, and 9, Table 1. This constancy is not achieved with the Type II formulation, Table 2. Moreover, the coefficients of *L* and *K* add up to close to 1, suggesting no economies of scale.

Regression equations 7 and 8, Table 1, produce similar results for *L* and *K* and also for *R* and *T*. But, when both *R* and *T* are introduced together in the regression equation 9, Table 1, neither regression coefficients are significant at the 5 percent two-tail level with 148 degrees of freedom. However, the performance of research and development is far better than that of time and, it seems, the success of the time variable is essentially due to its summary of the research and development variable. But when both are present, their intercorrelation reduces the level of significance of both, much more drastically in the case of time than research and development. The credibility of this interpretation of the role of *T* is enchanced by noting the erratic performance of *T* in other regressions, 4 and 5, Table 1, and regression 4, Table 2. It seems plausible, therefore, that the firm constants and the research and development variable summarizes the variables that are hoped to be captured by the conventional inclusion of *T* as a separate variable in the regression analysis of production functions.

Having analyzed the regression results, and on the basis of the above conclusion, regression 7 will be used to calculate marginal products of capital and research and development expenditures.

The elasticity of capital and research and development are respectively 0.16 and

0.11. The average real value added per firm per year was \$171.8 million, the average real gross capital per firm per year was \$407.6 million, while average real accumulated research and development expenditures per firm per year was \$35.2 million. An estimate of marginal product for capital is given by $0.16 \times 171.8/407.6 = \$0.09$ and that for research and development expenditures by $0.11 \times 171.8/35.2 = \$0.54$. That is, if the stock of capital increases by \$1.00 permanently, the value added increases by \$0.09 per year indefinitely. Similarly, \$0.54 worth of value added can be created per year indefinitely if the stock of research and development is increased by \$1.00 and the new level maintained permanently. This means that gross return on investment in research and development is 54 percent as compared with 9 percent for capital.

Without being over confident in the accuracy of the estimates, which for one thing are affected by the particular form of the production function used, the estimating procedure employed here has a few merits that may be pointed out briefly. First, in contrast to the tendency to find return to successful projects, a procedure which is on par with estimating returns to the acting profession by looking at the earnings of movie idols; my estimate reflects both successes and failures of research and development projects, which is as it should be. The value added increases due to successes, and research and development expenditures include both the cost of the projects that failed and those that bore fruits.

Second, the estimate is obtained by a joint estimation of the independent variables, in contrast to residual approach where what we cannot explain (the residual) is attributed to technology or research and development activity.

Third, typical returns to research and development are found to be several thousands percent, with the exception of a few studies that report them in the hundreds. Perhaps this fact has served the basis for some economists to justify their plea for more government subsidized research and development. As there is no entrance problem in conducting research, the persistance of the fantastic difference between returns to alternative investment cast doubt on the economic significance of the available estimates, regardless of their statistical significance. In contrast, the estimate derived in this paper is in the range of plausibility, especially when we note that the return is "social" rather than "private."

To the extent that increased productivity, caused by research and development, is partially passed on to consumers via lower prices for goods, my estimate will tend to reflect social returns. Lower prices will tend to decrease the firm's profits and the rate of return, but will not affect my measure, since real value added incorporates the benefits created but received by consumers through lower prices.

APPENDIX

Construction of the Variables[4]

1. *Value Added.* Since data on raw materials consumed were not available, the value added was estimated as wage bill+profit before taxes (adjusted by other income or expense not related to production; i.e., dividends received or paid)+interest paid+depreciation. The value added was then deflated by the firm's price index; the Bureau of Labor Statistics wholesale price indexes of subcategories of the chemical industries were weighted by the output mix of the firms by broad categories of products to obtain the firms' price index.

2. *Labor.* Total wages, salaries, and

---

[4] For a more detailed description of the variables, see my "The Economics of Research and Development," *op.cit.*

other compensations such as vacation, sick leave, insurance, retirement funds, etc., of employees were taken as the contribution of labor to production. The total wage bill was deflated by the firm's wage rate which was obtained by weighting the BLS index of average hourly earnings of subclasses in the total chemical industry by the same weights used in the construction of the price indexes.

3. *Capital.*

a) *Real Gross Plant and Equipment.* Data on gross plant and equipment at the beginning of the year, gross investment, and gross retirement are broken down by categories and types and are given in the *Annual Reports-Form 10-K* submitted to the Securities and Exchange Commission by publicly-owned firms. These reports date back to 1934. The gross plant and equipment at the beginning of 1934 was deflated by the 1929 GNP implicit price indexes of new construction and durable equipment, respectively. Yearly gross additions to plant and equipment were deflated by the price index of the year and added to the stock of the real plant and equipment of the initial year successively. It was assumed that the retired capital was generally the oldest. Therefore, the 1929 GNP implicit price deflators were used to deflate the retirements of the years 1934–46. Equipment retired in 1947–57 was assumed to have been purchased during 1940–47, so each of the two price indexes of this latter period were weighted by the relevant gross private domestic investment for 1940–47 to obtain the average deflator. This average weighted deflator was then used to deflate retirements of 1947–57.

b) *Other Real Capital.* This category of assets consists of cash, accounts receivable, government securities and inventories of the firm. This category of capital was deflated by the GNP implicit price deflator. Thus total real capital of the firms was equated to the total of net real plant and equipment and the other real capital.

4. *Research and Development.* The firms' research and development expenditures which include depreciation of laboratories and laboratory equipment were deflated by the GNP implicit price deflator.

Due to unavailability of earlier data for most firms, the accumulation of research and development starts from 1948.

5. *Time.* For each firm time takes the value of 1,2 . . . 10 for 1948, 1949 . . . 1957, respectively.

DISCUSSION

Jora R. Minasian: The four papers deal with "investment" with different points of departure. Eisner deals with the more general estimation problem encountered in investment functions; Scherer investigates performance of investment under different market structures; Mansfield explores the characteristics, costs, trade-offs of research and development, and the diffusion process of an innovation; and I have been concerned with estimating rate of return to research and development expenditures.

Due to an unfortunate mix-up in the mail, I shall not be able to provide a detailed technical comment on Eisner's paper. However, one general comment on the econometrician's frustration: it is not at all clear how much of the confusion is due to (*a*) measurement without theory and (*b*) theory without measurement.

Scherer tests a few alternative hypotheses, advanced by others, relating to the stability of investment and the structure of the industry. He finds that the variability of investment is correlated, among others, with concentration ratios, which is taken to indicate the market structure.

I have two comments to make: one dealing with the choice of the dependent variable and the other relating to the interpretation of the results.

Stability of investment—the dependent variable—is measured by the standard error of the estimate of a regression line of investment on time. The standard error is estimated from total industry data. The estimate is a function of the pattern of investment of both the "oligopolist" present in the industry (defined as industry with high concentration ratio) and of the investment behavior of the rest of the firms in the industry. In the extreme, the oligopolist may have a smooth investment trend line, but the residuals or the deviations from the industry investment trend line may arise from the fluctuating nature of the investment done by the rest of the firms constituting the industry.

The basic problem with the variable chosen and the test employed is that the hypotheses considered, whatever their merits, relate to the behavior of the firm, whether or not oligopolist, but the measured dependent variable does not. An appropriate dependent variable would be, for example, the standard error of the estimate derived from investment trend line of the firms that are included in the calculation of the concentration ratios. (I leave aside the dubious role of concentration ratios in implying different firm behavior.)

Given that the dependent variable is defined by the use of industry data, the result —the higher the concentration ratio, the less the stability of investment—has an interesting implication: the result is consistent with the hypothesis that the variability of investment is due to either entry into the industry and/or adjustment in production efforts of the firms already inside the industry. This may mean that there is more competition present in such industries than is usually assumed—the boat is being rocked.

Mansfield reports interesting findings. Collection of some relevant information is not an easy task, and I sympathize with him. Although some of his findings agree with my biases, I would like to make a few remarks to caution the reader.

The main results are derived from information collected by survey and opinion sampling. One has to be careful in concluding from such data. Specifically, information about risks, costs, and profitability are particularly suspect. One may ask, what is the incentive of the manager of an R and D laboratory? In justifying his demands for resources, should he claim high or low probability of success? Should he base the worth of the project on expected high or low profitability?

Studies of the predictability of costs, time, and success in military sphere indicate tremendous difference between expected and actual costs. Although Mansfield provides some information about the estimated and the actual probabilities attached to the success of the projects, we still need more information on the subject.

The least satisfactory state of affairs surrounds the information on profitability of the project. Not so accurate estimates are available for the projects even on *ex post*, let alone in an *ex ante* sense.

THEORY OF THE FIRM AND OF MARKET STRUCTURES

ADVERTISING AND THE ADVANTAGES OF SIZE*

By WILLIAM S. COMANOR, *Stanford University*
and
THOMAS A. WILSON, *University of Toronto*

This paper reports on some aspects of a continuing study of the economic effects of advertising. In a previous paper,[1] we studied the interindustry effects of advertising, and found that advertising has a substantial impact upon profit rates. In that paper, we emphasized the influence of product differentiation achieved via advertising upon barriers to the entry of new firms. As has been frequently noted, however, product differentiation can play a dual role. Not only may it affect the height of entry barriers but also it may influence competitive relationships among established firms. The present paper presents some evidence bearing upon the latter effect of product differentiation via advertising.

While advertising may be a strategic factor in the rivalry between large firms in imperfectly competitive markets, it may also confer advantages upon large firms as a group vis-à-vis their smaller rivals. Advertising and the associated product differentiation may influence the relationship between large firms and small by affecting both relative costs and relative prices. These are in fact closely tied effects, for smaller firms may have to spend considerably more per unit of output than their larger rivals to achieve effective price parity. This depends, of course, on the existence and relative importance of economies of scale in advertising. In this respect, both the size at which such advantages of scale are exhausted as well as the profit rate disadvantage associated with a lower volume of sales are important.

If each firm in an industry sold a similar menu of products, and if the prices set for these products were constant across firms, then differences in profit rates would reflect differences in unit costs. When products are differentiated, however, increased profits reflect not only lower costs but also higher prices. In such markets, an important function of advertising to the firm is to permit the achievement of price parity with highly advertised rival products. Because of these outlays, consumers will generally consider highly advertised products to be equal in "quality" to other major brands on the market. Where firm advertising outlays are low, on the other hand, firms are generally unable to reach effective price parity, for consumers will treat their products as substandard and their prices will be correspondingly lower. We have only to observe the striking price differences which exist among competing brands of aspirin, soap, or various cosmetics to recognize the significance of this factor. For this reason, where product differentiation is relatively great, differences in profit rates among firms will reflect differences in prices as well as in costs.

Where economies of scale in advertising exist, one might therefore encounter one or both of the following situations: (*a*) the smaller firms compete in the "branded" goods market at the expense of incurring

---

* The authors gratefully acknowledge the financial support of the Harvard Institute for Economic Research, the Brookings Institution, and Stanford University.

[1] W. S. Comanor and T. A. Wilson, "Advertising Market Structure and Performance," *Rev. of Econ. and Statis.*, Nov., 1967, pp. 423–40.

higher unit advertising costs; (*b*) the smaller firms may specialize in less differentiated standardized products that are priced below their highly advertised rivals.

To determine the extent of economies of scale in advertising, ideally one should examine the relationship between advertising inputs and outputs. From such an analysis, it would be possible to determine whether or not an increase in inputs led to a proportionately greater increase in outputs. This approach suffers, however, from the difficulties inherent in measuring the output of advertising expenditures. Since the "output" of large-scale advertising may be a higher price rather than lower unit costs, one cannot analyze this question directly by examining unit advertising costs in relation to firm size. Rather, the advantages of large firm size will be better detected through an analysis of the relationship between profit rates and firm size.[2]

Our objective is therefore to examine the net effect of advertising on profit rates for firms of different sizes within industries in order to obtain some evidence on whether advertising conveys advantages to large firms. We deal with this problem in two ways. First we examine the influence of these outlays on minimum efficient size of firm in relation to the market. Then we estimate the interindustry effects of advertising intensity upon profit rate differences between larger and smaller firms in an industry.

Estimates of Minimum Efficient Size of Firm

To examine the relationship between firm size and profit rates within industries, we assume, following earlier studies,[3] a functional relationship between rates of return and firm size which allows profit rates to increase with size, but at a decreasing rate such that profit rates approach an asymptote. Using this functional form, we can determine that level of firm size at which the asymptote is effectively reached. This provides an estimate of that firm size at which virtually all of the advantages of scale in terms of higher profit rates are exhausted. Note that we examine here the gross rather than the net effect of firm size. This is done because we are not concerned with the effect on profits of size per se as much as with the entire set of advantages which may be associated with firm size. This level of firm size will serve as an estimate of minimum efficient firm size[4] in each industry.

In obtaining estimates of minimum efficient firm size (MEF), we are concerned with the influence of scale on profitability and not with the plethora of other factors which influence the profit rates of individual firms. It is possible that the size of the most profitable firm in an industry may exceed the value estimated, but this would suggest only that other firms exist which are nearly as large and which have lower rates of return. For this reason, we obtain an estimate of MEF from the estimated relationship between firm size and profitability rather than directly from the size of the most profitable firm in the industry.

---

[2] It is important to note that economies of scale in advertising may result not only from quantity discounts in the sale of advertising messages, but also from the increased effectiveness of advertising messages as the number or quality of messages increases. The defense of a large market share may not require a proportionately large share of advertising messages. If, as Blank has suggested ("Television Advertising: The Great Discount Elusion, or Tony-Pandy Revisited," *J. of Bus.*, Jan., 1968, pp. 10–38), the former source of economies is relatively unimportant, our results would then relate predominantly to the latter source.

[3] See S. S. Alexander, "The Effect of Size of Manufacturing Corporations on the Distribution of Rate of Return," *Rev. of Econ. and Statis.*, Aug., 1949, pp. 229–35; H. O. Stekler, *Profitability and Size of Firm* (Berkeley, Inst. of Bus. and Econ. Res., Univ. of California, 1963); and Marshall Hall and Leonard Weiss, "Firm Size and Profitability," *Rev. of Econ. and Statis.*, Aug., 1967, pp. 319–31.

[4] While we use the term "minimum efficient firm size," we do not mean that such a size is necessary to obtain efficiency in any normative sense, but rather that it is the minimum size at which all of the private advantages of scale—whether social or antisocial—are attained.

Following our earlier study, the analysis is confined to forty-one consumer goods industries.[5] The measure of profit rates used is the ratio of profits after taxes to stockholders' equity, which is averaged within each asset size class in each industry, while size of firm is measured by the average value of assets per firm within the same size class. Each of the two variables is also averaged for the period 1954 through 1957, which covers a complete business cycle.

The empirical analysis consists of estimating for each industry a regression equation across size classes of the following form:

$$\pi = a + b\frac{1}{s} + c\frac{1}{s^2} + u \tag{1}$$

where π denotes average profit rates in the particular size class, s denotes average firm size, and u is an error term. Although fourteen asset size classes are included in the published IRS statistics, the bottom three classes were not used in the analysis since the profit rate statistics were considered especially unreliable. These classes include firms with total assets of less than $100,000 for which profit withdrawals in the form of executive salaries are likely to be important.[6] Although data for eleven size classes are available for most industries, a smaller number of observations was used in some cases because there are no firms in one or more size categories.

Equation (1) was not estimated directly because of the likelihood that the error terms are heteroscedastic. A previous study indicated that these error variances are approximately proportional to the reciprocal of firm size.[7] If the observations referred to individual firms, the appropriate correction would therefore be to multiply each observation by the square root of firm size. Since the observations refer to groups of firms by asset size class and the number of firms in each class is not constant across size classes, the appropriate correction factor is the square root of the total value of assets in the particular size class.[8]

---

[5] For a description of the consumer goods industries included and the definitions and sources of the variables used, see the appendix to our previous paper. (Comanor and Wilson, *op cit.*, pp. 438–40.) The various data for these industries are obtained at or aggregated to the IRS minor industry level, which is comparable, for the most part, to the SIC three-digit level. This classification scheme is not perfect for several reasons. First, some IRS industry groups contain more than one properly defined industry. Second, all of the activities of a diversified firm are allocated to the firm's most important industry. Finally, the effects of advertising upon profitability at different firm sizes should be related to the specific products or product groups of the firm, which may vary between firms in a given IRS industry. Such classification errors may distort the estimated intra-industry regression equations relating profit rates to firm size, and may therefore affect the estimates of MEF.

[6] See George J. Stigler, *Capital and Rates of Return in Manufacturing Industry* (Princeton Univ. Press), pp. 125–27. In our previous paper, and in the analysis of profit rate differences below, we excluded two additional size classes, confining the analysis to firms with assets above $500,000. To exclude these additional observations in the estimation of the equations relating profit rates to firm size would have cost two scarce degrees of freedom. Visual inspection of the profit rate data revealed that profit rates for the first three size groups were highly volatile, with a number of large positive and large negative observations. However, the profit rates of firms in the next two size classes did not show such volatility. In order to avoid reducing a small number of degrees of freedom even further, we decided to include these two size classes, at the possible cost of introducing some systematic error into the estimated coefficients.

[7] Hall and Weiss, *op. cit.*, pp. 323–24.

[8] If e denotes the error term in an equation analogous to (1) but which refers to individual firms, then

$$\sigma_e^2 = \frac{\sigma^2}{S}, \quad \text{and} \quad \sigma_u^2 = \frac{\sigma_e^2}{N},$$

As indicated above, we are concerned here not so much with the estimated regression coefficients as with the value of the function as it approaches the asymptotic value of a. Since the function never actually reaches this value, we evaluate it at a point which is arbitrarily close to the value of the asymptote. In the analysis below, we evaluate the function at 99 percent of a. This value is obtained by solving the following equation for S:

$$0.99\hat{a} = \hat{a} + \hat{b}\frac{1}{s} + \hat{c}\frac{1}{s^2} \tag{2}$$

where $\hat{a}$, $\hat{b}$, and $\hat{c}$ are the estimated regression coefficients.[9]

The resulting estimates of minimum efficient firm size are presented in Table 1. Note that estimates are given for only twenty-nine industries out of the forty-one included in the sample. In the remaining cases, the coefficient $\hat{b}$ had a positive sign so that the function approached the asymptote from above rather than from below, and therefore estimates of MEF could not be obtained. In these cases it appeared that the larger firms did not generally have higher profit rates. A comparative analysis of the characteristics of these industries in relation to the other twenty-nine is presented in the next section.

In addition to the estimates of MEF, Table 1 provides estimates of the standard errors of these statistics as well as estimates of minimum efficient plant size (MES) taken from our earlier study.[10] The average sizes of the leading four and the second four firms in each industry are also presented.

Before proceeding, it is necessary to consider one factor which may distort the estimates of minimum efficient firm size. The largest asset size class in the IRS statistics includes all firms with assets greater than $250 million. No account is taken of the possibility that larger firms within the upper size category may show higher profit rates than smaller firms within the same category. Distortions in the estimated nonlinear equation may be important where a substantial number of firms have assets exceeding $250 million and where substantial differences in size exist among these firms. In the motor vehicle industry, for example, much of the distribution of firm sizes is included in this single open-ended category. While this problem is most severe in that industry, it may be significant for other industries where the average size of the largest four firms exceeds $250 million. In these cases, some doubt must be cast on the estimates of MEF obtained.

As shown in Table 1, the standard errors of many of the estimated coefficients are relatively large. This is not surprising since the estimates are based on a maximum of eight degrees of freedom. At the same time, in twenty cases out of twenty-eight, the estimated coefficients do exceed their standard errors, while in seven industries the coefficients are greater than twice their standard errors.

As expected, minimum efficient firm size generally exceeds minimum efficient plant size, although there are six instances

---

where N is the number of firms in the size class. Therefore,

$$\sigma_u^2 = \frac{\sigma^2}{S \cdot N}.$$

Since, however, no information is available on individual firms, we assume that the size of each firm in a size class equals the class mean, and therefore,

$$\sigma_u^2 = \frac{\sigma^2}{S \cdot N} = \frac{\sigma^2}{\sum S},$$

where ΣS is total assets of all firms in the class.

[9] The largest positive root of the quadratic equation is taken to represent the appropriate economic magnitude.

[10] Comanor and Wilson, *op. cit.*, pp. 428–29.

TABLE 1

ESTIMATES OF MINIMUM EFFICIENT FIRM SIZE
(Value of Assets in Millions of Dollars)

Industry	Estimate of Minimum Efficient Firm Size	Standard Error of Estimate*	Estimate of Minimum Efficient Plant Size	Average Size of Firms	
				Top Four Firms	Second Four Firms
Soft drinks	14.2	19.5	0.8	38.4	11.9
Malt liquors	123.5	32.6	20.3	129.4	46.7
Distilled liquors	340.8	199.5	14.7	301.4	77.2
Dairy	18.9	14.7	2.1	342.4	52.7
Canning	64.6	23.1	6.7	190.6	79.2
Grain mill products	26.5	23.7	4.2	150.5	51.4
Cereals	43.6	41.4	36.5	100.9	4.6
Bakery products	7.7	10.3	2.6	154.0	46.6
Sugar	78.1	47.6	31.0	110.3	42.4
Confectionery	61.6	23.4	10.3	69.2	21.1
Cigars	41.8	30.9	5.4	38.2	8.5
Cigarettes	39.3	23.5	99.6	558.5	174.8
Knit goods	24.3	34.1	2.0	46.9	14.9
Men's clothing	15.7	6.4	1.7	41.2	15.4
Women's clothing	6.8	14.1	0.5	13.9	7.3
Furs	37.9	3.9	0.2	0.9	0.6
Furniture	34.9	11.0	2.3	49.5	19.2
Periodicals	14.1	52.4	22.4	93.0	33.3
Drugs	21.0	17.0	46.5	131.4	85.1
Soaps	11.2	58.1	18.7	323.2	8.0
Paints	35.5	23.0	8.7	235.6	30.7
Perfumes	52.0	67.0	10.9	32.6	14.3
Footwear	11.2	15.1	2.0	110.0	15.4
Handtools	32.7	27.0	16.6	65.2	21.9
Electrical appliances	55.0	38.8	21.2	71.2	11.5
Radio, T.V. and phonograph	43.7	36.2	20.4	319.4	111.8
Motor vehicles	60.6	175.5	375.0	2,482.1	1,545.4
Instruments	20.6	19.3	28.4	520.7	67.7
Jewelry (precious metal)	31.6	9.8	4.6	31.1	9.3

* Since the estimate of minimum efficient firm size is obtained from a nonlinear relationship with the estimated regression coefficients, its standard error cannot be obtained directly. The standard errors presented in this table are derived from the first term of Taylor expansion of the nonlinear equation around the true value and are therefore only approximate values.

where the opposite relationship is found. No economic significance should be attributed to the latter, since the differences are never statistically significant, and the estimates of minimum efficient scale of plant are themselves subject to error. Note also that among these six industries, four contained firms sufficiently large that the average size of the leading four firms exceeded $250 million, and therefore differences in size within the open-ended size category may have distorted the estimates.

A more interesting comparison is with the actual sizes of the largest four and next largest four firms in the industry. In twenty-four cases out of twenty-nine, the average size of the leading four firms exceeds the size which exhausts all of the advantages of scale, and in thirteen of these, the differences are statistically significant. In four out of the five industries where the estimates of MEF exceed the average size of the largest four firms, on the other hand, the differences are not significant. In many industries, therefore, the average size of the largest four firms is clearly greater than necessary to realize all private economies. These statistics confirm the conclusions of earlier studies which indicate that the sizes reached by the largest firms in many industries cannot be explained on the basis of economies of scale alone.

When comparisons are drawn between the estimates of minimum efficient firm size and the average size of the second largest four firms in an industry, different conclusions are reached. In twenty-six of the twenty-nine industries, the average size of these firms exceeds estimated minimum efficient plant scale. In seventeen industries, however, the mean size of the second largest group of firms lies below MEF. Thus it appears that while the size of firms in the second largest group is typically sufficient to exploit all economies of scale in production, these firms, unlike their larger rivals, are frequently not large enough to have exhausted the full advantages of size. This suggests that the top four firms may frequently have a competitive advantage over the next largest group, which is confirmed by the analysis of profit rate differences below.

Advertising Intensity and Minimum Efficient Firm Size

In this section, we examine the influence of advertising intensity and production economies of scale on the proportion of the market accounted for by a firm of minimum efficient firm size. Various interindustry regression equations are estimated in which the dependent variables are the estimates of MEF expressed as a proportion of the market.[11]

A dummy variable to distinguish the three local or regional industries in the sample—soft drinks, dairy, and bakery products—was introduced into the regression equations to correct for the use of the size of the national market in the denominator of the dependent variable. Since the ratio of MEF to market would be substantially higher in these cases if appropriate market definitions were used, the coefficient for this variable should be negative.

In specifying the regression equations, we do not assume that the effects of advertising and production economies are necessarily additive. Where economies of scale at the plant level are quite large, the optimal firm size required to realize these economies may be sufficient to realize any economies of scale in advertising. At the same time, firms which are sufficiently large to realize in full any advertising scale economies are likely to be sufficiently large to concentrate their production in plants of efficient scale. On these grounds, an interaction variable—the product of the advertising-sales ratio and the ratio of minimum efficient plant scale to market—was introduced. Both linear and logarithmic variants of the basic equation were estimated.

When all four independent variables are introduced together as explanatory variables, none of the regression coefficients, except that of the regional industry dummy variable, is statistically significant in either the linear or logarithmic versions due to the collinearity between them. Hence the empirical results presented in Table 2 involve alternative equations with one or two of the variables omitted. In equations (1) and (2), the interaction variable is introduced alone in the linear and logarithmic versions, respectively. In both

---

[11] The estimates presented in Table 1, which are expressed in the volume of assets, are multiplied by the appropriate industry ratio of sales to assets and then divided by the value of average industry sales in the particular industry for the four-year time period. In order to avoid using estimates which lie beyond the range of observable sizes, the values of MEF in the regression analysis are constrained to be less than or equal to the average size of the largest four firms in the industry. For those five industries where the estimates exceeded this level, the constraining value was used to construct the dependent variable.

TABLE 2

MULTIPLE REGRESSION EQUATIONS EXPLAINING RATIO OF MINIMUM EFFICIENT FIRM SIZE TO MARKET

Dependent Variable	Constant	Regression Coefficients	R^2	Corrected R^2
1. MEF/Mkt =	$0.041^{\dagger}$ (0.011)	$+5.307^{*}\,(Ad/S \cdot MES/Mkt) - 0.048\,REG$ (2.521) (0.031)	0.19	0.13
2. $Ln\,MEF/Mkt$ =	$-2.530^{\dagger}$ (0.497)	$-0.072^{*}\,(Ln\,Ad/S \cdot Ln\,MES/Mkt) - 1.497^{*}\,REG$ (0.032) (0.655)	0.23*	0.17*
3. MEF/Mkt =	0.026 (0.016)	$+0.199\,Ad/S + 0.466^{*}\,MES/Mkt - 0.070^{*}\,REG$ (0.283) (0.263) (0.035)	0.21	0.11
4. $Ln\,MEF/Mkt$ =	-1.591^{*} (0.914)	$+0.203\,Ln Ad/S + 0.323\,Ln MES/Mkt - 1.558^{*}\,REG$ (0.199) (0.190) (0.682)	0.24	0.15

Notation:

MEF/Mkt	Ratio of estimated minimum efficient firm size to market.
MES/Mkt	Ratio of estimated minimum efficient plant size to market.
Ad/S	Advertising-Sales Ratio
REG	Regional industry dummy variable which has value of one if industry is regional, zero otherwise.

* Indicates statistical significance at the 5 percent level.
† Indicates statistical significance at the 1 percent level.
Figures in parentheses are the estimated standard errors of the coefficients.

cases, its coefficient is statistically significant but the coefficient of multiple determination is significant only in the logarithmic equation. In equations (3) and (4), the two component variables are introduced independently but without the interaction term. In these equations the advertising coefficients do not exceed their standard errors although the ratio of MES to market is significant in one of the two equations. However, the coefficients of multiple determination are insignificant in both equations. In all four equations, the regional industry dummy variable has a significant coefficient with the expected negative sign.

The negative sign on the interaction variable in the logarithmic equation indicates that the partial effects of both the ratio of MES to market and of advertising intensity are positive since both variables are necessarily less than one and therefore have negative logarithms. This negative coefficient also indicates that, as expected, the effect of advertising upon the ratio of MEF to market diminishes as the relative importance of plant economies increases, and vice versa.[12] To examine the size of this effect and the extent to which it diminishes, we can compute the partial elasticities of MEF/Market with respect to the advertising-sales ratio at alternative levels of MES/Market.[13] Some typical values are given below:

RATIO OF MES TO MARKET	ELASTICITY OF MEF/MARKET WITH RESPECT TO ADVERTISING-SALES RATIO
0.01	0.33
0.03	0.25
0.05	0.22
0.10	0.17

---

[12] Note that the linear equation with the interaction variable alone (equation (1) in Table 2) implies the opposite effect since the relevant coefficient is positive. It is, of course, impossible under this specification to have partial positive effects of advertising intensity which also diminish with the ratio of MES to market. This may account for the lower R^2 obtained with this specification.

[13] These computations are readily made by noting that the relevant

$$\text{elasticity} = \frac{\partial\, ln\, MEF/Mkt}{\partial\, ln\, Ad/Sales} = -0.0718 \cdot ln\, MES/Mkt.$$

These results suggest that advertising outlays interact with scale economies at the plant level to influence the minimum size of an efficient firm. Particularly where scale economies are small in relation to the market, heavy advertising outlays may increase the minimum share of the market accounted for by an efficient firm.

Before concluding this section, it is useful to examine the twelve industries which have been excluded from the regression equations. These industries are those for which the hypothesized positive relationship between profit rates and firm size did not hold. Table 3 presents some pertinent structural characteristics of the excluded industries. The ratio of MES to market is somewhat higher in these industries, although the difference is not significant. On the other hand, these industries show lower advertising-sales ratios, on the average, than those industries which were included in the regression equations, which suggests that this factor may have been partially responsible for the inverse relationship between profit rates and firm size.[14]

TABLE 3

CHARACTERISTICS OF EXCLUDED* INDUSTRIES

Industry	Advertising Sales Ratio	Plant Economies of Scale
Wines	5.2%	4.2%
Meat	0.6	0.7
Carpets	2.0	7.7
Hats	2.2	10.4
Millinery	0.8	0.4
Screens and blinds	1.6	3.3
Books	2.4	1.8
Tires and tubes	1.4	5.0
Household and service machinery	1.9	6.4
Motorcycles and bicycles	1.1	12.5
Clocks and watches	5.6	8.3
Costume jewelry	4.0	0.7
Average—12 excluded industries	2.4	5.1
Average—29 included industries	3.7	4.5

* These industries are those for which estimates of minimum efficient firm size could not be obtained and which are excluded from the regression equations presented in Table 2 above.

Profit Rate Differences between Large and Small Firms

As noted above, we are concerned with the determinants not only of minimum efficient size of firm but also of the extent of the disadvantages suffered by firms which are smaller than this size. In this section, we deal with the latter problem by examining differences in profit rates between firms at different scales of output. This variable is used to represent the net effect of differences in both prices and costs.

Profit rate differences are defined, for each industry in our sample, between various sets of "large" and "small" firms, and these differences are then related to two industry characteristics—the relative importance of economies of scale at the plant level and the relative importance of advertising.

The effect of economies of large-scale plants upon profit differences should depend on the extent to which the production of firms in both size groups is concentrated in plants of suboptimal scale, upon the average size of such suboptimal plants, and upon the shape of the production cost curve below minimum efficient plant scale. We assume that all firms in the larger size group will have fully exploited such economies, and that the disadvantages of the smaller group of firms is related to their mean size in relation to minimum efficient plant size. We do not use the ratio of MES to market because, although used in our previous paper to measure the effects of scale economies on average profit rates in an industry, it should not be directly related to profit rate differences between two groups of firms within an industry.

---

[14] We tested whether the mean advertising-sales ratio in "excluded" industries is significantly greater than the mean ratio in the "included" industries. The t statistic fell slightly below the critical value at the 5 percent confidence level (1.56 as compared with a critical level of 1.68) and therefore we were not able to reject the null hypothesis at that confidence level. The null hypothesis could be rejected, however, at the 10 percent level.

The effect of advertising upon profit rate differences should depend not only on the average advertising intensity in the industry but also upon the relative differences in size between the two groups of firms. This effect is measured, therefore, by the product of the advertising-sales ratio and the logarithm of the ratio of the average size of the larger firms to the average size of their smaller rivals.[15]

To test these hypotheses, we estimate interindustry regression equations of the following form:

$$(3) \quad PD = a + b\, Ln(SL/MES) + c[(Ad/S)\cdot Ln(SH/SL)] + v,$$

where

PD is the difference between the mean profit rate of the larger firms and the mean profit ratio of the smaller firms,

SL/MES is the ratio of the mean size of the smaller group of firms to the scale of a minimum efficient plant, where both variables are measured by the volume of sales,

Ad/S is the advertising-sales ratio for the industry,

SH/SL is the ratio of mean sales of larger group to mean sales of the smaller group of firms, and

v is an error term.

If advertising intensity and plant scale economies were the only factors affecting the profit rate advantage of the larger firms, the functional form of the equation would imply zero intercept. The intercept is included in order to allow for the possible net effect of other factors which create advantages of size. In addition to the general equation, equations with one of the variables and/or the intercept suppressed are also presented.

Profit rate differences between three alternative sets of large and small firms are examined. In the first set, we examine only firms above MES and analyze the difference between the average profit rates of the top four firms and that of the remainder of the firms above minimum efficient plant scale. While it is unlikely that profit differences between these groups are related to economies of scale at the plant level, this variable is introduced into some of the regression equations to allow for possible errors of measurement in the estimates of MES.

The results are presented in Tables 4 and 5. We find that, on the average, the top four firms enjoy an average profit rate advantage over the remaining firms above MES of 1.6 percentage points. This difference, which amounts to about a 20 percent advantage, is statistically significant at the 5 percent level. The multiple regression equations suggest that these differences are attributable largely to the effects of advertising adjusted for difference in average firm sizes between the two groups. Since the intercept is not significant in any equation, this implies that, once the effects of product differentiation via advertising are taken into account, there is no net advantage, on the average, to the largest firms. As expected, the measure of plant scale disadvantages of smaller firms is insignificant, which simply reflects the fact that all of the smaller firms included in this particular comparison lie above minimum efficient plant scale.

In the second analysis, firms in each industry were divided into those above and

---

[15] The logarithm of this ratio and the logarithm of the ratio of SL to MES are used in equation (3) because we assume that the effect upon profit rates of a given absolute increase would diminish as size increases.

TABLE 4

TESTS OF SIGNIFICANCE OF MEAN PROFIT RATE DIFFERENCES

Profit Rate Difference* Between:	Mean Difference†	Standard Deviation of Difference	N	t Test
Four largest firms and other firms above MES.	.0160	.0408	35‡	2.32‖
Firms above MES and firms below MES.....	.0248	.0406	38§	3.76**
Four largest firms and all other firms.........	.0252	.0380	41	4.24**

* Firms with total assets below $500,000 are excluded from all calculations.

† All averages are weighted averages of relevant asset size class means; the weights are proportional to total equity for the asset size class.

‡ See note * for Table 5 below.

§ See note † for Table 5 below.

‖ Indicates statistical significance at 5 percent level.

** Indicates statistical significance at 1 percent level.

below estimated minimum efficient plant scale. Here, of course, we would expect to find that the effects of economies of scale at the plant level are more important. This expectation is confirmed by the empirical results.

As shown in Table 4, firms above MES have a mean profit rate which is 2.5 percentage points higher than firms below MES, an advantage of roughly one-third. This difference is statistically significant at the 1 percent level. However, unlike the difference between the two groups above MES examined in the previous paragraph, advertising plays no role in explaining this profit rate differential. In none of

TABLE 5

INTERINDUSTRY REGRESSION EQUATIONS PREDICTING PROFIT DIFFERENCES BETWEEN LARGER AND SMALLER FIRMS

*A. Equations predicting the profit rate difference between the largest four firms and all other firms above MES**

Equation	Constant	Effects of		R^2
		Plant Economies	Advertising Intensity	
		$\left(Ln\frac{SL1}{MES}\right)$	$\left(\frac{Ad}{S}\cdot Ln\frac{SH1}{SL1}\right)$	
A. 1	−.00133 (.0143)	.000247 (.0113)	.520§ (.237)	.131
A. 2	−.00155 (.0102)		.520§ (.233)	.131§
A. 3		.00097 (.0081)	.509‖ (.205)	.131
A. 4			.493‖ (.146)	.131§

TABLE 5 (*Continued*)

B. *Equations predicting the profit rate difference between firms above MES and the firms below MES*†

Equation	Constant	Effects of		R^2
		Plant Economies	Advertising Intensity	
		$\left(Ln\frac{SL2}{MES}\right)$	$\left(\frac{Ad}{S}\cdot Ln\frac{SH2}{SL2}\right)$	
B. 1	−.00777 (.0134)	.0406# (.0133)	−.0233 (.0805)	.213§
B. 2		.0349# (.0089)	−.0393 (.0749)	.205§
B. 3	−.00910 (.0124)	.0398# (.0128)		.211#
B. 4		.0315# (.00607)		.199#

C. *Equations predicting the profit rate difference between the largest four firms and all other firms*‡

Equation	Constant	Effects of		R^2
		Plant Economies	Advertising Intensity	
		$\left(Ln\frac{SL2}{MES}\right)$	$\left(\frac{Ad}{S}\cdot Ln\frac{SH1}{SL3}\right)$	
C. 1	−.0155 (.0173)	.0209 (.0139)	.566§ (.235)	.187§

* Six industries were deleted because there were no firms above MES that were not also in the top four, so these regressions were fitted to thirty-five observations.

† Three industries were deleted because there were no firms below MES that were above the size cutoff of $500,000 in assets, so that these regressions were fitted to thirty-eight observations.

‡ Both sets of industries mentioned in footnotes * and † were deleted, so that this regression was fitted to thirty-two observations.

Definition of Variables
$SH1$ = Mean sales for top four firms
$SL1$ = Mean sales for remainder of firms above MES
$SH2$ = Mean sales for all firms above MES
$SL2$ = Mean sales for firms below MES (with assets above $500,000)
$SL3$ = Mean sales for all firms except the top four
MES = Estimated minimum efficient scale of plant
Ad/S = Advertising sales ratio for the industry

§ Indicates statistical significance at the 5 percent level.
# Indicates statistical significance at the 1 percent level.

CALCUTTA UNIVERSITY CENTRAL LIBRARY

the regressions is the scaled advertising-sales ratio significant, and the observed profit rate advantage of firms above MES is attributable largely to the extent to which the smaller group of firms lies below MES. The variable measuring this effect is significant at the 1 percent level in all equations. The constant term as well as the scaled advertising variable is insignificant, indicating that once the effect of technical economies is taken into account, firms above MES have no net profit rate advantage. These findings also provide additional support for the validity of our estimates of scale economies at the plant level.

Since the advantages of size in relation to advertising intensity appear to be particularly important for the largest firms in an industry, an additional analysis was carried out to examine profit rate differences between the four largest firms and all others both below and above MES. Because the previous analysis demonstrated that firms above estimated MES suffered no disadvantage in the scale of production facilities, the ratio of the mean size of firms below MES to MES was used to measure the interindustry effects of plant scale economies upon the profit rate advantages of the top four firms. The measure of advertising intensity used in this analysis is the advertising-sales ratio scaled by the logarithm of the ratio of the mean size of the top four firms to the mean size of all other firms.

The estimated equation is presented in the third part of Table 5. The scaled advertising intensity variable is statistically significant at the 5 percent level, whereas relative production scale economies are only significant at the 10 percent level. The insignificance of the constant term suggests that the net effect, on the average, of other factors upon these profit rate differences is unimportant.[16]

Taken together, these three sets of regression results indicate that product differentiation via advertising confers an advantage upon the very largest firms in an industry and, as would be expected, that firms below minimum efficient plant scale suffer a disadvantage attributable to economies of scale in production.

Conclusion

The interindustry analyses of profit rate differences and minimum efficient firm size provide some empirical support for the hypothesis that there are advantages to size which are related to the level of advertising intensity in an industry. These results suggest that there are net advantages to large firms—beyond those attributable to economies of large-scale plants—in industries where product differentiation via advertising is an important dimension of industry structure.

---

[16] This is not to argue that the partial effects of other factors are unimportant, since the effects of separate factors may offset one another.

SURVIVAL-ABILITY AS A TEST OF EFFICIENCY

By Joe S. Bain
University of California, Berkeley

At least since the days of the Temporary National Economic Committee before the second World War, economists have shown recurrent interest in determining empirically, in each of many industries, the comparative efficiency of firms of different sizes. As we progress from the smallest toward the largest firm in any industry, do unit costs of production, distribution, and perhaps also sales promotion tend to vary systematically with firm size?

One type of study has attacked this question directly by attempting to measure the long-run average cost or scale curve of "the firm" in each of a number of industries. Two things may be noted about such measurements. First, they generally attempt to abstract from interfirm cost differences due to differences in productive technique, in money prices paid for productive factors, and in rate of utilization of plant. Second, they generally admit sales promotion costs into a scale curve only provisionally and with qualms, because of a priori and empirical indications that unique and reversible functional relationships between firm size and unit sales promotion costs—available alike to all firms—probably do not exist in most industries.

Three general findings concerning scale curves have emerged from this sort of study. First, in the usual industry, economies of large scale are realized as the firm expands from miniscule size to some identifiable minimal optimal scale, but not beyond that scale. Second, unit costs tend to remain constant at the optimal level as the firm grows still larger, and diseconomies of very large scale ordinarily do not seem to be encountered. Third, most of the scale economies observed are economies of the large-scale plant. Although these findings apply specifically to scale curves measuring only production and distribution costs, they are roughly sustained when sales promotion costs are included, except that in some product-differentiated industries smaller firms seem to encounter diseconomies of very large-scale sales promotion.

Three other connected findings also deserve mention. First, in the great majority of industries, a firm of minimum optimal scale, or of a smaller scale not involving significantly higher unit costs, would supply an appreciable but distinctly minor fraction of the output of its industry. Second, in most industries, the largest firms are from several to many times larger than minimum optimal scale. Finally—and for purposes of the ensuing argument most important—a significant minor proportion of industry output—in one sample of industries from 10 to 30 percent—is generally supplied by a fringe of small firms that suffer significantly from diseconomies of small scale (primarily of small-scale plant). And these fringes of inefficient one-plant firms appear to survive over considerable periods of time—perhaps indefinitely.

This last is a tentative finding of some theoretical interest and policy importance. Yet before it was developed, George Stigler, writing in 1949, advanced a proposition that might appear to contradict the finding and to deprecate the research effort underlying it. Let me quote him as follows:[1]

---

[1] George J. Stigler, "Monopoly and Oligopoly by Merger," *A.E.R.*, May, 1950, pp. 23–34.

> The comparative private costs of firms of various sizes can be measured in only one way: by ascertaining whether firms of the various sizes can survive in the industry. Survival is the only test of a firm's ability to cope with all the problems: buying inputs, soothing laborers, finding customers, introducing new products and techniques, coping with fluctuations, evading regulations, etc. A cross-sectional study of the costs of inputs per unit of output in a given period measures only one facet of the firm's efficiency and yields no conclusion on efficiency in the large. . . . If a firm of a given size survives, we may infer that its costs are equal to those of other sizes of firms. . . . [However,] the equality of private costs carries no implication that social costs of firms of different sizes are equal.

He also implicitly suggested as a tentative indicator of "survival" of firms of a given size class whether or not they were maintaining or increasing their relative numbers in the industry.

Given this seeming disagreement between empirical findings and a priori doctrine, two related questions arise. First, is there a genuine contradiction or conflict between the two? Yes or maybe. There is definitely a conflict between the criteria of efficiency reflected respectively in the findings and in Stigler's doctrine. The findings refer to the comparative efficiency of firms of different size as measured approximately in terms of social costs (neglecting externalities generated by firms), whereas Stigler clearly espouses private costs as a criterion of efficiency. This choice of criterion, moreover, is not simply a neutral matter of taste. Conclusions referring to interfirm differences in social costs have a clear normative content; those emphasizing only private costs have at best a descriptive one. From a normative standpoint, moreover, one might object to Stigler's terminology when he asserts that a firm's advantage derived from such things as buying factors at lower prices and evading regulations improve its "efficiency in the large."

On the other hand, the empirical findings in question and Stigler's proposition are not of necessity contradictory on a technical level. Because they rest on significantly different definitions of cost, both could be correct simultaneously, at some level of probability. But if there is, as seems plausible, a fairly high degree of correlation between social and private costs, the presence of some contradiction seems on balance likely.

This leads us to our second question: In his proposition was Stigler probably correct on his own ground—that where only private costs matter? Unfortunately, we cannot dismiss the matter by observing that his proposition is tautologically true and not subject to confrontation with evidence, simply because he has arbitrarily defined efficiency as survival-ability. He has precluded us from this by holding that all surviving firms in an industry, regardless of size, have equal private average costs. To deal with the issue, we must examine the sources of the survival power of socially inefficient firms on which Stigler expressly or implicitly had to rest his case, and judge to what extent these sources are likely to give all surviving firms equality in private costs. We thus inquire generally as follows. Why can smaller firms in an industry that are comparatively inefficient in terms of social unit costs survive persistently or for appreciable periods of time?

Before turning directly to the question, let us note that there are some instances in which small survivors are actually not surviving in the same industry as the larger firms with which they are compared, because of misleading Census or other groupings. They are supplying essentially different products than those large firms, and products which are distant substitutes to any common group of buyers. The simultaneous survival, for example, of Ferrari and Aston-Martin, on the one hand, and General Motors and Fiat, on the other, has about as much to do with comparative efficiency as the simultaneous survival of General Foods and National Steel. The first and second pairs of auto firms mentioned have different production functions, are on different scale curves, and are essentially noncompeting. A comparison of their unit costs (as well as of their survival-ability) would be meaningless.

In the same vein, let us also note that small surviving firms in some industries, in spite of evident diseconomies of small-scale production, may not actually be less efficient, socially or privately, when all costs are taken into account. The best examples of this phenomenon are perhaps found in some industries where the larger firms are localized to a few geographical points (possibly to minimize assembly costs of raw materials) and where the product has high outshipment costs relative to selling price. Then, small firms may be able to locate in and supply geographically remote local markets, offsetting diseconomies of small-scale production with reduced costs of delivery. The American steel industry, for one, appears to fit this mold fairly well, as does the petroleum refining industry. In such cases, of course, the smaller firms are probably surviving because of equivalent social efficiency, and not in spite of comparative inefficiency.

To these examples we might add (with some misgivings about including selling costs in measures of efficiency) examples of small surviving firms that systematically offset diseconomies of small-scale production with sales promotion costs per unit of output lower than those of their larger rivals. This strategy, which may put the smaller firms in some sense on the same level of social efficiency as that of the larger ones, is not conducive to large or growing market shares for the small firms. But it may suffice to maintain them with small capacities that are suboptimal from the standpoint of production costs. I am hard put to identify any "pure" examples of the phenomenon in question, but it would appear to have been alloyed with other considerations; for example, in the shoe industry and the petroleum refining industry.

With these and similar false or misleading examples of survival with inferior social efficiency out of the way, let us turn to some valid grounds for such survival. The following explanations of the phenomenon seem particularly important in the light of empirical observation:

1. *Living under a Price Umbrella.* Consider industries with market structures featuring a rather concentrated oligopolistic core of a few large firms, together with a competitive fringe of smaller firms that are clearly less efficient by either social or private criteria. In a number of these it has been observed that the dominant firms establish prices noticeably in excess of their own average costs, but roughly equal to the average costs of their smaller rivals. This policy, in addition to yielding excess profits to the larger firms, allows the smaller firms to survive with low or normal profits, and perhaps keeps them hungry enough to discourage their expansion. Such a structural situation, featuring the persistent survival of a group of relatively inefficient small firms, may be reasonably stable over time because of product-differentiation disadvantages of the small firms and of other intra-industry barriers to their expansion.

Supposing that it is, it is important to note that the large and small firms are generally—unless other factors intrude—surviving at distinctly different levels of affluence (e.g., different private rates of excess profits on sales). And equal survival-ability at quite different levels of affluence definitely does not itself imply equal private costs, nor equal social costs. This phenomenon will probably not explain a majority of the cases we are interested in, but it does not seem to be especially rare.

In some product-differentiated industries, of course, it is possible that some small sellers with both social and private inefficiency do not need a price umbrella in order to survive. Without the benefit of having the product prices of their major rivals held above average costs, they can or could command minor market shares at superior selling prices sufficiently high to offset their diseconomies of small scale. To the extent that such sellers could survive in this way, they might be considered to have at least adequate allocational efficiency, by virtue of supplying the wants of buyer minorities with distinctive products which would otherwise be unavailable. But "pure" cases of this phenomenon, unalloyed with price umbrellas, are hard to find.

2. *Surviving with the Support of Large Rivals.* In some very highly concentrated oligopolies, there has been evidence that the dominant firms have pursued policies designed to support the survival or avert the disappearance of their few relatively inefficient small rivals, generally in order to preserve the appearance of competition and thus avoid antitrust charges of joint monopolization. One such policy of course involves holding a price umbrella. But others include supplying the small rivals with materials on favored terms, giving them access to controlled distributive outlets, etc. This phenomenon is unquestionably rarer than the first mentioned, but it has been evident, for example, in the automobile industry, wherein the Big Two or Three evidently made efforts to keep Kaiser from going under after its entry to the industry, and where the failure of American Motors would be a minor disaster to the principal firms. Small firm survival secured by this route may reduce small firm private costs, but seems in general unlikely to equalize them with those of larger firms.

3. *Securing Hired Productive Factors at Lower Money Prices.* There are evidently a number of industries in which small firms, that are definitely less efficient than larger rivals in terms of real factor inputs, partly or wholly offset their virtual cost disadvantage by securing labor and other factors at lower prices than the large firms do. In some cases this is done by exploiting the geographical and possible occupational immobility of the labor force in a declining area. (It is not a sign of an offsetting social efficiency unless labor mobility is strictly not a viable alternative.) In some cases it is accomplished through favored treatment by unions or through locating so as to use nonunion labor. In all these cases—and kindred ones as they may affect other productive factors—the attainment of or approach to equality in private costs by socially inefficient small firms is possible. Of course, this fails to erase the fact that these firms have found a way to survive in spite of inferior real efficiency. The shoe industry of New England appears to have offered a good case in point.

4. *Entrepreneurial Absorption of Losses.* In a number of industries—generally of only moderate concentration but very easy entry—we have observed that a persistent survival of numerous small firms with definite diseconomies of small scale has been made possible mainly by one fact. That is, owner-entrepreneurs are persistently willing to accept profits yielding subnormal returns or no returns on their investments, or imputed wages below their market value in other employment, or both. In some instances, the same inefficient firms survive for long periods of time. In others, such firms periodically go

bankrupt, but their plants remain in operation as another wave of entrepreneurial cannon fodder assumes ownership with the same odds against it. In either event, we may find persistent survival or its equivalent in spite of clear social inefficiency, based on a reduction of private costs. This source of survival power may frequently be coupled with monetary advantages in securing hired factors.

This phenomenon has persisted for a long time in the New England region of the shoe industry. It has been continually observed in gasoline retailing, where the owner-operators of small-volume service stations stay in business in return for total monthly personal earnings approximately equal to the market rental values of the self-owned lands on which their stations are situated.

5. *Favored Treatment or Subsidy by Governmental Bodies.* It is a well-known fact that small corporations and other small firms receive favored treatment in the matter of federal income taxes and that this may condone survival by reducing private costs in spite of comparative inefficiency. It is also known that numerous local or state governmental bodies have offered tax remissions to new firms in order to attract industry. Such remissions may of course favor large and small firms alike, but there is some probability that they may serve to keep inefficiently small firms alive.

6. *Holding a Market Through Unfair Trade Practices.* Among the ways of surviving other than being as efficient as your competitors are those of deception of consumers, design piracy, commercial bribery, and so forth. The considerable evidence of such practices encouraged the passage of Section 5 of the Federal Trade Commission Act, and in particular of the Wheeler-Lea amendment thereto, in 1937. An aim of this legislation was to establish a "plane of fair competition" on which efficient firms would prosper and given which less efficient firms could not prosper or survive through various forms of "cheating."

Yet unfair competition does persist, and though large as well as small firms may play the game, it appears in a number of industries that small firms of inferior social efficiency do survive mainly by engaging in deception of consumers and other unfair trade practices. Instances in which this is true are perhaps not extremely numerous, but segments of the pharmaceutical and cosmetic industries would seem to offer fair examples. In some cases, such use of unfair trade practices may support small-firm survival by allowing them to reduce their private costs to a level competitive with those of larger rivals. In others, it may simply enable them to elevate their selling prices, neither private nor social costs being reduced.

The preceding list of sources of survival power of socially inefficient firms is of course not exhaustive, but are there any generally important sources we have omitted? In his 1949 article, Stigler mentioned as advantages of small firms which might let them secure competitive private costs in spite of inefficient scales the following: buying inputs, soothing laborers, evading regulations, finding customers, introducing new products and techniques, and coping with fluctuations. The first three of these are essentially covered under points 3 and 6 above. It is not apparent that small firms systematically or usually have significant off-setting advantages in the latter three respects, and we will not consider them here.

We may now return to the magnum question which led us into an exploration of the main sources of survival-ability of socially inefficient small firms. That is, to what extent was Stigler correct in holding that small firms with survival power have private costs (though not necessarily social costs) that are equal to those of larger surviving firms. In view of the preceding argument, it seems that he was possibly correct so far as the bases of small-firm survival-ability—other than equal social costs—are lower money factor prices, entrepreneurial absorption of losses, favored treatment by governments, some unfair trade practices, and some sorts of artificial respiration administered by large rivals. Yet it is not apparent on a priori grounds or from observation that these sources of survival-ability must operate to give smaller firms full equality in private costs. This might be required in some models of atomistic competition, but not in actual markets.

Correspondingly, he was incorrect to the extent that the survival-ability of small firms depends on their living under price umbrellas held by large rivals or on some varieties of unfair trade practices. To this extent, his original proposition should be amended.

How serious is this amendment likely to be? The answer lies in factual information that has not been adequately developed and analyzed, but I would hazard a few guesses on the basis of accumulated observation. First, within the competitive fringes of firms of inefficiently small scale, which in most industries appear to supply from 10 to 30 percent of industry outputs, an appreciable minority do not deserve to be designated as socially or privately inefficient. This is either because they have been grouped for comparison with larger firms in different theoretical industries, or because some of their compensating real advantages have been neglected. Second, a clear majority of these fringe firms are socially inefficient, all things considered. Third, within this latter group, survival-ability depends wholly or in part on things other than equality in private costs—mainly on living under price umbrellas—in a substantial minority of cases. A close scrutiny of a sample of twenty manufacturing industries with which I have previously worked suggests that price umbrellas contribute strongly to small-firm survival-ability in about 40 percent of the cases. The indicated amendment to Stigler's proposition thus appears to be more than trivial.

In conclusion, what of the so-called "survivor" technique? This of course embodies a sort of reverse-English application of Stigler's proposition by trying to identify which size classes of plants or firms in Census industries have minimal private costs by determining which size classes survive the best, in terms of holding or increasing either their comparative populations or their shares of industry value added over selected time intervals. (It should be noted that even when the focus is nominally on optimal plant sizes, firm sizes are also involved because in the bulk of industries plants in the smaller size classes are predominantly held by one-plant firms.) I will not attempt an appraisal and critique of this novel sort of numerical exercise here, deferring to W. G. Shepherd's masterful review of issues in his recent article in the *Southern Economic Journal.*[2] But I will note that the economic arguments outlined above suggest that the survivor technique among other things rests on shaky conceptual foundations, as applied either to firms or to plants. All things considered, survival-ability is an uncertain and perhaps treacherous indicator of even private efficiency.

---

[2] William G. Shepherd, "What Does the Survivor Technique Show About Economies of Scale?" *So. Econ. J.*, July, 1967, pp. 113–22.

ALLOCATIVE EFFICIENCY AND THE LIMITS OF ANTITRUST*

By OLIVER E. WILLIAMSON
University of Pennsylvania

Allocative efficiency, like any simple criterion applied to the formulation of antitrust policy, is subject to limitations. For one thing, the criterion that is proposed here is not value free. Absent compensation, which for a variety of reasons is rarely feasible, acceptance of such a criterion implies a willingness to make (what some will regard as arbitrary) redistributional judgments. For another, the operationality of efficiency as an antitrust standard has been contested. Other objections might also be raised (for example, the merits of a piecemeal approach such as that employed here can be disputed on "second best" grounds), but these are beyond the scope of the paper.

An interpretation of the allocative efficiency criterion is provided in Section I. This begins with an examination of what would appear to be an unnecessarily restrictive specification of the allocative efficiency standard. I then attempt progressively to relax certain of these restrictions, and go on to advocate that a middle-range version of the allocative efficiency criterion become the principal basis for formulating antitrust policy and enforcing the Sherman and Clayton Acts. Although occasional exceptions to this standard may be warranted, these are apt to be both few in number and limited in scope: proposed exceptions can often be appropriately digested within the suggested allocative efficiency framework.

The judicial limits of the middle-range allocative efficiency standard are assessed in Section II. A program for gradually introducing the criterion into judicial proceedings—at a minimum, to recognize the standard in principle—is proposed. Commonly neglected efficiency attributes of conglomerate organization, the innovation process, and product variety are briefly examined in Section III.

I. *Allocative Efficiency Interpreted*

Three separate approaches to allocative efficiency analysis can be distinguished. One is concerned with the immediate allocative efficiency consequences of a proposed change, and may or may not include a sensitivity to trade-offs between market power and economies effects. The second is the middle-range standard in which both trade-offs and anticipated future consequences are allowed for. The third involves an extension of allocative efficiency analysis to embrace "fairness" within the framework of demoralization cost analysis. Each is considered in order. The emphasis throughout is on mergers, but much of the argument generalizes easily.

A. *The Immediate Standard.* Robert Bork, in a paper presented at these meetings two years ago, advanced what is essentially an allocative efficiency standard for antitrust [11]. Although I am broadly in accord with his position, I will argue that the standard he proposed is unnecessarily narrow[1] and that a more general formulation of allocative efficiency is to be preferred.

The criterion that he advances, which

---

* Research on this paper has been supported by grants from the National Science Foundation and the Brookings Institution. Helpful comments on earlier drafts by Almarin Phillips, Richard Posner, and Donald Turner are gratefully acknowledged. Responsibility for the views expressed, however, is my own.

[1] He develops the theme further as part of his review of the legislative intent of the Sherman Act in [10]. Also see [6] [7] [8] [9].

he characterizes as a consumer welfare standard, is defined as "minimizing restrictions of output and permitting efficiency, however gained, to have its way" [11, p. 242]. Elsewhere he expresses this objective as "the maximization of wealth or consumer want satisfaction" and indicates that "this requires the courts to distinguish between agreements on activities that increase wealth through efficiency and those that decrease it through restriction of output" [10, p. 7]. In applying this allocative efficiency criterion to mergers, he expresses skepticism about the notion of incipiency [6, pp. 368–69], is evidently unconcerned with the question of timing, indicates that technical progress is irrelevant to an efficiency evaluation [8, p. 832] [11, p. 251], and regards "fairness" as outside the main tradition [8, p. 838–39]. The last issue is treated in Part C below; here I consider the partitioning, incipiency-timing, and progressiveness aspects of the argument.[2]

The suggestion that activities can be exhaustively divided into those that produce economies on the one hand and lead to market power on the other would appear to be an oversimplification. It would also misread Bork's intent. It is not so much the absence of mixed cases as it is the difficulties of measuring efficiencies that causes him to shun the trade-off issue [9, p. 390].[3] Inasmuch, however, as it is easy to conceive, in principle at least, of mixed cases,[4] the allocative efficiency standard may require that trade-offs be recognized (even if only immediate first consequences are considered) if incompleteness in the analysis is to be avoided.

Although Bork does not dismiss the matter of incipiency, he omits altogether a discussion of timing. An emphasis on immediate in relation to eventual effects is thereby suggested. Again the risk is that consequences will be incompletely characterized. Thus, while the evaluation of immediate and eventual effects may often be identical, counterexamples are easily constructed. As I have argued elsewhere, to disallow a merger need not reflect a decision to forego the realization of combinatorial economies altogether; ordinarily it is not a question of whether economies will be realized but when and with what market power effects [35, p. 27]. For this it is necessary to consider prospective as well as immediate consequences, perhaps in the local rule-making sense discussed below (see Part B).

Bork's treatment of progressiveness likewise suggests an immediacy emphasis. He takes the position that while efficiency is essential to consumer welfare, "the propriety of 'progressiveness' as an antitrust criterion is not obvious. . . . Progress will occur . . . at that rate which consumers choose by the degree to which they make it profitable to engage in the activity of producing progress" [10, p. 251]. But plainly, progressiveness is an aspect of dynamic efficiency in a cost reduction sense [36, pp. 29–31] [37], and product variety can be interpreted in efficiency terms as well (see Section III). Inasmuch as progressiveness may not be independent of market structure (there are both theoretical and empirical [29] reasons for believing otherwise), an effort to merge progressiveness with the efficiency standard would seem indicated. Presently, however, neither the arguments nor the evidence relating market structure to progressiveness is unmixed. Provisionally, it may only be judicious to withhold judgment on this dimension. But, as the evidence accumulates, sharper definition is ordinarily to be expected; an operational integration of progressiveness within the efficiency standard might, therefore, eventually be achieved.

---

[2] Two other issues that are raised by this statement of objectives are the social-private cost distinction and the suggestion that wealth and consumer welfare are equivalent. Although the phrase that economies "however gained" shall have their way easily lends itself to a private cost interpretation, this presumably was unintended. Equivalence with respect to wealth and consumer welfare may obtain by definition, but this nevertheless requires (explicitly or implicitly) that distributional considerations be faced. Bork's insistence that redistribution be ruled out as an objective of antitrust [8, pp. 837, 839] might be read as a suggestion that the welfare function be specialized such that benefits and costs are weighted equally whatever their incidence.

[3] Although Bork appears mainly to prefer to partition economic events into those which have either efficiency or market power consequences [7, p. 41] [8, p. 830] [10, pp. 7, 36] [11, p. 249], he nevertheless recognizes that mixed cases can develop [8, p. 832, 838] [9, p. 390].

[4] If one assumes that combinatorial economies are not unusual among firms with only modest market shares, and if the Supreme Court is correct in detecting market power effects (present or potential) in recent merger cases involving firms in relatively unconcentrated markets, then the conjunction of market power and economies may indeed be common. But this argument, it should be noted, is merely one of existence rather than degree. Unless, however, the predictable market power and economies effects both exceed threshold proportions, the trade-off question (operationally) is never reached.

Overall, therefore, while I would agree with Bork that an allocative efficiency standard has much merit, freeing it up to deal with the mixed cases, explicit concern over process (incipiency and timing), and an integration of progressiveness with efficiency analysis would each appear to be desirable. This entails no significant alteration of the criterion; a simple reinterpretation so as to embrace (at least aspects of) each of the foregoing is easily supplied—as is illustrated, for example, in [36] [37]. The immediate standard must thus give way to a middle-range approach.

B. *A Middle-Range Standard.* My purpose here is to examine a standard which makes allowance for concurrent market power and economies consequences and is explicitly concerned over eventual effects (incipiency, timing, progressiveness). The question of whether antitrust should be actively concerned with income distribution considerations is also considered.

The partial equilibrium apparatus that I would propose for purposes of examining the trade-off question is one in which the welfare function is expressed as $W = (TR + S) - (TC - R)$, where TR refers to total revenue, S to consumers surplus, TC to total cost, and R to intramarginal rent.[5] The terms in the first set of parentheses reflect the social benefits associated with the activity in question, while the terms in the second (under appropriate restrictions) reflect social costs. The allocative efficiency consequences of any merger that increases both efficiency and market power can be evaluated only by estimating net effects.

Extending the above trade-off analysis from a treatment of immediate to include eventual effects will ordinarily require the analyst to consider industry as well as local consequences. Incipiency thus involves a judgment of the prospective market power consequences of allowing a merger in comparison with those that will obtain if economies are realized by internal expansion instead. What potential rivalry between the firms in question (the issue in the El Paso case [34]) or what subsequent combinations (as raised in the Brown Shoe case [13]) can reasonably be expected? The latter part of this question is concerned with the probability that the merger under examination signals a trend. Relevant to an evaluation of this is whether the proposed merger has significant characteristics that distinguish it from industry conditions more generally.

A judgment with respect to the timing of a merger is industry related in that "optimal" timing is apt to be much affected by the rate of growth of market demand: *ceteris paribus,* the greater the rate of growth of market demand, the easier it is to achieve the requisite expansion by internal means and (assuming that changes in market structure, once introduced, tend to persist) the less compelling the case for combination. The rate and direction of inventive activities (dynamic aspects of efficiency) similarly involve the analyst in industry issues—albeit difficult as these are to evaluate at this time.[6] To the extent, however, that nontrivial industry-related consequences predictably obtain as a result of incipiency, timing, or progressiveness, these presumably should be given effect by augmenting the basic tradeoff model accordingly. General guidelines, such as those recently issued by the Department of Justice [33], may have to be supplemented by industry studies, such as those favored by the Federal Trade Commission (cf., [21]).

---

[5] Implicitly, this is the welfare function that I employ in [36]. The absence of externalities is ordinarily, but not necessarily, presumed.

Income distribution is the next issue to be treated. Consider in this connection J. de V. Graaff's contention that "by far the simplest way of securing the distribution of wealth we desire is *through the price system*. . . . Much of orthodox welfare theory lacks realism precisely because it assumes that the desired distribution of wealth has already been attained (and is somehow maintained), and then proceeds to regard the price system as a highly specialized resource-allocating mechanism which exercises no influence whatever on the distribution of wealth" [23, p. 155]. The allocative efficiency standard proposed here violates both parts of Graaff's argument: antitrust policy is not regarded as an efficient instrument for the redistribution of income, while the question of a "correct" distribution of income is assumed away.

Necessarily, for every allocative efficiency judgment which has distributional consequences, there exists an implicit, if not an explicit, distributional weighting. Typically the benefits and costs are weighted equally, "to whomsoever they may accrue." This would be appropriate if one were to assume that the political process had already "solved" the distributional problem so that, subject to the conditions that the changes proposed are small, neutrality at the margin obtains.[7] Even if this does not hold exactly, an "imaginary" neutrality assumption might be supported by the argument that antitrust is poorly suited to advance redistributional objectives. Macroeconomic policy instruments (taxes, transfers, expenditures) with which to correct distributional conditions are not only available but are superior to the use of antitrust for this purpose. Typically these instruments involve the promotion of regional development, subsidizing a group that is involved in a transitional adjustment, alleviating conditions of poverty, etc., and are applied on a broad scale. Antitrust, by contrast, is concerned with industry conditions of a much more local variety. Also, as Bork points out, the prohibitory nature of the antitrust laws makes antitrust a poor instrument by which to pursue distributional goals of a general sort: "There seems no rational design in entrusting to judges . . . [an income distribution] role under a statute which haphazardly limits their effectiveness to those values which by chance happen to be capable of implementation through a wholly prohibitive law" [8, p. 837].

Except, therefore, as an allocative efficiency criterion can be shown to have nontrivial exploitative consequences on an identifiable subset of society, antitrust might best be enforced by suppressing redistributional considerations. Moreover, where systematic exploitation of this sort exists, the indicated remedy is to provide a legislative exception rather than a judicial correction. To involve the courts in such merit choices is inadvisable: "There can be few more intensely political determinations and few for which the judicial process is less suited" [8, p. 839].

---

[6] For a brief review of the evidence, see [36, pp. 29-30].

[7] One might wish to stipulate that the distributional changes effected will subsequently be taxed according to the rules under which neutrality was realized.

If, realistically, a legislative exception is not feasible, one might introduce a threshold stipulation instead: unless the expected net allocative gain (expressed, say, as a ratio of real cost savings to deadweight losses) exceeds some specified value, an economies defense will not be entertained. Although this would be a crude adjustment—and in some cases might be counterproductive even from an income distribution point of view—it would tend to assure that potentially unfavorable distributional consequences would not easily be sacrificed for only modest efficiency gains. At the same time, an efficiency orientation would be preserved and active involvement by the courts in the income redistribution process would be avoided.

C. *The Criterion Extended.* The argument above is that an allocative efficiency standard limited to immediate consequences is too restrictive but that an attempt to embrace income distribution objectives is too ambitious. A middle-range standard is therefore proposed. This does not, however, exhaust the power of allocative efficiency analysis. It can be extended to make allowance for what Frank Michelman has referred to as "demoralization costs" [31]. One form that these can take is that of social discontent. It was presumably problems of this sort that stimulated Senator Sherman's observation that "the popular mind is agitated with problems that may disturb social order, and among them none is more threatening than the inequality of condition, of wealth, and opportunity that has grown within a single generation out of the concentration of capital into vast combinations to control production and trade and to break down competition" [15, p. 2460]. A similar type of concern is suggested by Frank Knight's observation that "the more vital problems are not problems of economy, but of maintaining social unity in the face of economic interests. . . . The social problem is to preserve respect for the rules, and to make such rules as result in the best game for all, players and spectators" [27, pp. xxxii-xxxiii]. While conditions have changed since Sherman spoke in 1890 and Knight reflected on the circumstances of society in 1933, the views expressed have a certain element of timelessness; fairness as well as efficiency needs to be considered.

Social discontent is not, however, the only form that demoralization costs can take. Adaptive responses that take the form of asset reallocations are also possible. These reallocations are induced by the failure of compensation to be paid when investment expectations that had been regarded as secure are upset by actions of the state. Owners and other interested investors adapt by allocating fewer resources to activities which, despite attractive returns, are apt subsequently also to be upset. The principal relevance of this aspect of demoralization cost theory to antitrust enforcement is in dissolution proceedings. Thus whereas others have emphasized the equity aspect of dissolution or divestiture on "hapless stockholders and employees" [16], the demoralization cost argument directs attention to the efficiency (resource allocation) implications of such a policy.

As I have shown elsewhere [38], the conventional externalities framework can easily be extended to include adaptive responses of both types: social discontent and asset reallocation costs are, in principle, appropriately included in a full-blown allocative efficiency analysis. In terms of the welfare function given above, social costs are understated when these are expressed as direct pecuniary costs less intramarginal rent if secondary responses predictably obtain.[8]

The question then is what is the relation of this analysis to the matter of equity and fairness. Will one, on extended allocative efficiency grounds, generally come out in favor of policies which are more commonly (and often fervently) supported on grounds of fairness? The answer that both Michelman [31, p. 1223] and I [38] supply is a qualified "yes," but the reasons for this can scarcely be reproduced within the limits of this paper. Suffice it to observe that those who contend that a principal function of antitrust is to forestall social discontent may find that allocative efficiency can often be invoked to support the usual equity arguments on which this claim rests.

Such an extended version of allocative efficiency analysis is not, however, well suited to case-by-case application. What is needed is a policy designed to deal with the class of behavior for which secondary responses can be reasonably predicted. Rule making rather than case-by-case review is therefore indicated.[9] Presumably such rules would take the form of exceptions to the usual middle-range framework.

Finally, consider the extraeconomic problems of sociopolitical significance that are raised by large absolute size. Here a problem of incommensurability with the allocative efficiency criterion is encountered. The necessary judgment, again, is one better made by the legislature than the courts. Possibly, as Carl Kaysen has suggested, this would take the form of a prohibition of expansion by merger of the largest 50 or 100 corporations [25, p. 37]. Assuming that exceptions on such grounds do not bulk large as a fraction of the total, a middle-range allocative efficiency standard would remain the principal basis for enforcing the antitrust statutes.

II. *The Limits of Antitrust*

The middle-range allocative efficiency standard proposed would allow an economies defense for a merger that has both efficiency and market power consequences provided that (1) the relevant industry-related consequences were given effect and (2) legislative exceptions did not prohibit the merger in question. Our purpose here is to examine the judicial limitations that this approach is subject to and to suggest a program for gradually introducing the criterion into the judicial process.

There is general abhorrence by legal specialists in antitrust to the suggestion that economies be an admissible defense for a merger having mixed effects. Donald Turner's affirmative views on economies, mainly in the context of conglomerate mergers [32], are clearly an exception. Bork prefers rules (market share limits) to trade-off analysis on the grounds that the measurement of efficiencies, "for all practical purposes, is impossible" [9, p. 390]. The recently issued "Antitrust Guidelines" of the Department of Justice disallow an economies defense for all but "exceptional circumstances," which are left undefined [33]. The most thorough examination of this question, in what is probably the controlling article, is the 1960 treatment by Derek Bok [5]. Although Bok acknowledges that an economies defense deserves consideration in principle, he is apprehensive of its operational consequences. He proposes that only a failing company defense be allowed; even this is subject to reservations.

---

[8] Note that social discontent is separable into two types: "Aggressive" adaptations take the form of disruptive acts against society in response to what is regarded as the willful choice to exploit a group. "Despondency" is social discontent of an extreme sort. Productive activity on the part of losers and their sympathizers may be substantially reduced, and responsibility for their care turned over to the state.

[9] Case-by-case review may nevertheless be necessary to build up initial experience with the phenomenon in question before the rule is specified.

Bok takes the position that exceptions to a market share rule should be rejected, "however attractive they may be in theory, until there is reasonable assurance that they can improve the accuracy of our decisions in practice without imposing prohibitive losses in terms of clarity and simplicity of administration" [5, p. 287]. Economies, as he sees it, do not qualify. For one thing, Congress appeared not to be seriously concerned with economies in amending the Clayton Act [5, p. 318]. For another, the subject economies can frequently be realized through internal expansion [5, p. 319]. But the fundamental objection appears to be that a consideration of economies, or other "economic factors," would upset the legal process. The reasons for this are several.

First, there is a suggestion, perhaps unintended, that economies are only one of several proposed economic factors, and he discovers no apparent agreement among economists over which are most significant [5, pp. 270, 290–92]. Second, a merger may have mixed consequences, and there is no evident way of netting out the results, if indeed the effects can be measured [5, p. 290]. Third, even if the inclusion of economies (or other factors) led to a more rational result when applied to some mergers, holding out the possibility of such a defense might impose additional costs and lead to less satisfactory results in other proceedings [5, p. 294]. Finally, there is the possibility of overloading the judicial system; procedures that are designed to appear rational may, if the necessary processing capacity is absent, yield dysfunctional results [5, pp. 296–98].

Whether Congress was genuinely or only apparently unconcerned with economies would seem to be a legitimate issue. Relevant in this connection are the following questions: What was the fact framework within which the legislation was enacted? What implications should be drawn from the provision for a failing company defense? Were the trade-offs between allocative efficiency and other goals evident?[10]

Much of the congressional concern with the apparently mushrooming merger movement of the 1940's was based on a Federal Trade Commission report which, as Bok makes clear, was a seriously misleading document [5, pp. 231–34]. Had the post-1940 merger activity been characterized more accurately, the *Congressional Record* might read differently. But courts cannot revise premises, however incorrect, of congressional action. Nevertheless, taking the record as given, one is entitled to inquire what is to be implied from congressional provision for a failing company defense. Bok argues that congressional sensitivity to economies is not to be inferred. He suggests instead that the failing company defense "stemmed from a legislative concern over the various interests involved in the life of a failing enterprise. Creditors, owners, employees—all could have an interest in avoiding a total collapse or in realizing as high a selling price as possible" [5, p. 340].

---

[10] The argument that follows in the text essentially accepts the framework that Bok employs in evaluating congressional intent with respect to economies. One could, however, shift the grounds for establishing congressional intent altogether: since Congress did not disallow an economies defense in Section 7, such a defense is automatically to be presumed. Thus one might take the position that although the Sherman Act did not provide for an economies defense, the courts had effectively read such a defense into the statute (see for example Judge Wyzanski's interpretation in United shoe of Judge Hand's opinion in Alcoa). Had Congress, which was legislating against a history that included the recent Alcoa opinion, intended that an economies defense be unavailable for Section 7 cases, it would have made this explicit. My point in the text, however, is different: even neglecting this history, congressional insensitivity to an economies defense is not established by reference to congressional provision of a failing company defense.

Possibly congressional interest in these affected groups should be interpreted exclusively in distributional rather than efficiency terms. But ordinarily a concern for distributional values requires (at least rough) advance knowledge of the composition of the affected groups. The creditors, owners, and employees in firms of size sufficient to have market power consequences (which is the relevant subset, since otherwise antitrust considerations are not even raised) do not obviously qualify for special distributional relief. It seems not implausible, however, to impute to Congress sensitivity (albeit unarticulated) to efficiency effects. For one thing, the time horizon and, hence, the utilization and maintenance of resources by firms on the verge of bankruptcy are apt to be suboptimal. In addition, if aversion to risk predictably influences the allocative decisions of the creditors, owners, and employees in the affected firms, allowing a defense which limits the expected degree of loss will simultaneously have allocative consequences. The failing company defense thus admits of at least a limited variety of economies interpretation.[11] Perhaps, lacking a trade-off framework, this is as far as Congress would be expected to go. But given a trade-off framework, such as that developed in [36] and discussed above, we might push further. What are the effects of a broader economies defense on the legal process? Consider each of the four objections raised by Bok.

To recapitulate, the issues are these: an unstructured multiplicity of economic factors, absence of a trade-off framework, excessive administrative costs, and an overloading of the judicial system. Any suggestions that allocative efficiency is merely one of several competing economic factors, none of which have preferential standing, would appear to miss the "first principles" character of the efficiency goal. Efficient resource allocation, where this is broadly rather than narrowly conceived, is essential to the long-run vitality of the economy. Harry Johnson expresses it as follows [24, p. 69]:

> Economic theorizing, research, and policy discussion have tended to be excessively concerned with "macro-economic" . . . to the neglect of the "micro-economic" problems of efficient resource allocation, whose solutions are likely over the long run to be more important to the achievement of a highly productive and rapidly growing economy. In a fully employed economy, the main concern of economic theory is with these "micro-economic" problems of allocation of scarce resources among competing uses.

More general recognition of this "main concern" would presumably help to overcome the obscurant treatment frequently accorded to economies in the continuing dialogue on antitrust policy.

Assuming that allocative efficiency is accorded preferential standing, the trade-off question needs next to be faced. The approach developed in [36] and elaborated in Section I above is expressly concerned with this issue. Assuming that, in principle, this has been dealt with acceptably (including the treatment of exceptions and incommensurabilities), the objections to an economies defense turn on implementation rather than theory.

---

[11] Still an additional interpretation is that Congress was concerned with neither distributional nor allocative considerations but with "fairness" instead. Although operational content is difficult to provide for the notion of fairness, Michelman's examination of this issue [31] led him to conclude that the criteria for judging fairness are not apt to be significantly different from those invoked under the extended allocative efficiency standard discussed in Section I-C above.

What then of the measurement problem and of the disruptive effects that allowing an economies defense would have on legal proceedings? Although the measurement problem is a serious one, the difficulties can easily be exaggerated. These can be mitigated at the outset by requiring that claims of economies pass threshold specifications before such will even be entertained [36, p. 24]. Moreover, once an economies defense is admitted in principle, incentives are set in motion to sharpen up the specification of economies of various sorts. To dismiss an economies defense on the basis of the prevailing state of the art is to employ an unacceptably narrow horizon.

Whether the judicial process will be overloaded by allowing an economies defense depends partly on the characteristics of the program by which a change of this sort is implemented. Considering the present state of enforcement of the antimerger statute, merely a reorientation of attitudes regarding economies would have merit—and this could be obtained without admitting an economies defense quite generally. As things stand now, we observe the regrettable condition in which a company proposing a merger, an apparent effect of which is to realize economies, consciously suppresses the economies aspect lest it be used affirmatively by the government to attack the merger [12, p. 93]. Moreover, as the Brown Shoe case suggested, the concern may not be wholly idle speculation [13, p. 344].[12] Lacking legitimate standing for economies, we should not be surprised, perhaps, by statements such as that of the Federal Trade Commission indicating that "the necessary proof of violation of the statute consists of types of evidence showing that the acquiring firm possesses significant power in some markets *or* that its over-all organization gives it a decisive advantage in efficiency over its smaller rivals" [20, p. 1084]—a position which Turner characterizes as both bad law and bad economics [31, p. 1324].

Indicative, possibly, of changing attitudes by the government on economies is a statement in the government's brief in the Procter & Gamble case (which was prepared jointly by the Department of Justice and the Federal Trade Commission): "in general, advantages afforded by a merger which reflect simply greater efficiency ought not be a basis for holding the merger illegal; efficiency is, after all, a prime goal of antitrust" [18, p. 47]. The Supreme Court has yet to adopt this or similar language, however; equivalent language is also absent in the *Merger Guidelines* [33]. At best, economies would appear to enjoy a shaky legitimacy—possibly insufficient to withstand a change of enforcement officials. If retrogression were to occur, Bok's basic contention that "unless we can be certain of the capacity of our legal system to absorb new doctrine, our attempts to introduce it will only be more ludicrous in failure and more costly in execution" [5, p. 228] does not obviously hold. Judicial ease accomplished at the expense of perverse application of the law comes at a high cost. The consistent application of an irrational standard is not obviously optimal.

A four-stage program to implement an economies defense is proposed. At a minimum, the antitrust agencies and the courts should explicitly recognize the merits of an economies defense in principle, even though they may disallow it as a practical defense. This would at least forestall perverse treatment of economies before the law. The effects of even such a modest change might well be considerable.

---

[12] See Bork and Bowman [6, p. 373] and Turner [32, p. 1324] for an interpretation of the language of the Court on this question.

As Kaysen has pointed out, "policy change comes about, in large part, by the way in which the enforcing agencies select cases and frame issues for courts and commissions to decide" [26, p. 85].

Second, although an economies defense would be disallowed, it could be introduced for "explanatory completeness."[13] This would hopefully encourage companies to reveal the efficiency consequences of mergers more completely, so that the importance of allowing an economies defense could be better assessed. The legal process would be protected, and an idea of the magnitude of the implicit trade-offs might begin to emerge.

A more ambitious undertaking, but one which would still provide substantial protection to the legal process, would be to allow an economies defense for certain types of mergers. Justice Harlan has suggested that an economies defense should be available for conglomerate or product-extension mergers [17, pp. 1240–41]. As experience is accumulated, this might be eventually extended to include other types of mergers. With respect to vertical and horizontal mergers, for example, *Merger Guidelines* could be rewritten to acknowledge a band of uncertainty within which an economies defense would be available: above a specified set of market share values, a merger would be disallowed; below a second set of values, a merger would be permitted; within the range between, an economies defense, subject to an appropriate set of threshold stipulations [36], would be entertained. Proceeding in a gradualist way of this sort would appear to conform to the objective proposed by Bok [5, pp. 348–49]:

> While the basic values and objectives of Section 7 should presumably remain unchanged by court or agency, the methods and standards used to achieve these ends must evolve to keep pace with developments in our knowledge concerning the nature and effects of mergers. . . . The issue . . . is fundamentally one of timing; the question is whether *at present* we stand to gain from rules which seek to capture the subtler images of reality evoked by the complexities of economic theory.

Finally, an economies defense (appropriately qualified) might be admitted quite generally. Even, however, if this last stage is never reached, the process is worth setting in motion. Stage one involves merely an immediate expression of positive regard for economies by the antitrust agencies and the courts—a step which is unambiguously desirable. There are clear benefits and no evident costs.[14]

III. *Applications of Efficiency Analysis*

The antitrust question addressed in Section I is that of specifying an appropriate criterion. Section II deals with implementation. The concern here is with providing antitrust with a relevant world view. Obviously all are essential to an effective antitrust program.

Broadly my concern is with characterizing the enterprise system and the economy of which it is a part. The proposition that I wish to advance is that one of the most remarkable attributes of American capitalism is its adaptive capacity to devise efficient and viable organizational forms in the pursuit of private gain. A similar argument, expressed in a somewhat more qualified way, may also hold for product variety. To characterize the system in static, historical, or traditional terms may be to miss much of what accounts for its most significant achievements. To be relevant, therefore, antitrust may need periodically an updated interpretation of the phenomena with which it is concerned.

---

[13] As Bork points out, the courts have, occasionally, permitted this [9, p. 390, fn. 40].

[14] The sequential procedure proposed here is superior to the approach I suggested previously, where I advocated that the antitrust agencies perform a trade-off test themselves before challenging a merger [36, p. 34]. The earlier proposal spares the courts of the burden of treating the economies issue, but introduces what is probably an unacceptable degree of discretion into the administrative process.

Space considerations permit me merely to identify, without developing, three areas in which antitrust analysis may require rethinking. These are: the efficiency consequences of conglomerate organization; a systems approach to technical progress; and a "modern" characterization of product variety.

A. *Conglomerate Organization.* The first hypothesis that I would tentatively advance (which I explore more thoroughly elsewhere [39] [40]) is that the conglomerate organization—or, more generally, that the multidivision form—often permits the realization of substantial managerial economies. Typically such economies are admitted in principle but dismissed as *de minimis.* I suggest, however, that the multidivision form, regarded as a control device, has what are frequently superior properties. Both the suboptimization (by which I mean a technical variety of control loss [35]) and the subgoal pursuit (a motivational phenomenon) which are observed in the unitary form organization are apt to be attenuated. If this characterization of the multidivision form is substantially accurate, expressed misgivings over the efficiency consequences of the conglomerate form of organization [4] [18] need presumably to be qualified.

Lest this be misinterpreted, I am not suggesting that simply because the multidivision form has the capacity for superior control that improved control will necessarily obtain from divisionalization in any particular case. The manner in which the parts are administered is critical [39] [40]. Moreover, even where such control advantages are observed or can be expected, there may be countervailing considerations. The existence of such economies does not, therefore, place such a firm beyond the reach of antitrust. The point, simply, is that the managerial economies aspect of the multidivision form would appear to warrant greater weight than has been accorded to it previously.

B. *Technical Progress.* A related, if somewhat more conjectural, hypothesis is the following: an efficient procedure by which to introduce new products is for the initial development and market testing to be performed by small firms (perhaps new entrants) in an industry, the successful developments then to be acquired, possibly through merger, for subsequent marketing by a large, multidivision firm. The reasoning is this. Although the large, multidivision firm has apparent control advantages for dealing with routinized activities, it may be ill-suited to handle high-risk inventive activity. As with any large organization, it experiences difficulty in admitting error and its large resources permit it to persist beyond judicious cutoff limits. Its bureaucratic structure may also induce risk-aversion within the individual parts, as alleged by Norman Berg [3]. Although error admission may likewise be difficult in the small firm, the latitude needed to delay response is lacking. Moreover, the small firm may be less inhibited in assuming risk; venture capital is often earmarked for small enterprises which (because of investor inability to select among the investment projects in the large firm and appropriate the undiluted gains of high-risk activities) is unavailable to the large firm. Conceivably, therefore, small specialized organizations that are quickly responsive to market results may possess advantages in the new product development and testing areas. Once market success is reasonably assured, the scale and control economies of operating within the multidivision form may, unless the firm can easily achieve the requisite size and variety itself, dictate a transfer. Although the argument oversimplifies considerably [29], it is perhaps sufficient to suggest that aspects of the innovation process can usefully be regarded in system rather than individual firm terms. (The optimum system need not include among its parts any individually "optimal" firms.) If correct, parts of the recent discussion of innovation and firm size may require reworking. Also, the problem of maintaining (in a moving equilibrium sense) an optimum industry structure, especially assuring the requisite ease of entry for new product innovators, would need to be faced.

C. *Product Variety.* The animus with which product variety, including especially "product differentiation," is commonly treated appears to be due to the inadequacies of traditional theory, paternalism on the part of the analyst, or observed redundancy in the product offerings. The first needs to be remedied, the value judgment implicit in the second should be recognized, and a scientific basis for characterizing redundancy would appear to be indicated. The first and third of these require a richer allocative efficiency framework. Although its operational properties need more fully to be worked out, the skeleton of a promising framework has recently been provided by Gary Becker [2], Kelvin Lancaster [28], and William Baumol [1].

"Characteristics," not goods, give rise to utility. What needs to be evaluated in judging the social benefits of a good that offers a new combination of characteristics (i.e., a differentiated commodity) is how it affects the efficiency frontier. This may frequently demand a deeper level inquiry than the traditional antitrust approach has required if the real consequences of differentiation in any particular case are to be discovered. The point is scarcely new; it was made almost twenty years ago by Edward Chamberlin [14]. But at that time a promising analytical basis for dealing with differentiation was lacking.

IV. *Conclusions*

Edward Mason has observed that "there is an element of faith in the proposition that maintaining competition substantially improves the efficiency of resource use" [30, p. xx]. This would appear to be an accurate characterization of the prevailing uncertainties within which antitrust presently operates. It will doubtlessly continue unless antitrust can be provided with a basis for proceeding in a more self-confident way. Lacking a coherent analytical framework that can be progressively articulated, economists are apt to find, as Joel Dirlam and Alfred Kahn have suggested, "that the appropriate question for economists to ask about antitrust policy is not whether this is the most efficient way of structuring or reorganizing the economy, but the inverted one: Does antitrust seriously interfere with the requirements of efficiency?" [17, p. vii.]

It is perhaps natural to believe that restricting options results in a loss of power. Were the antitrust agencies to adopt an analytical standard of the sort proposed here, activities that are now subject to prosecution would be foreclosed. Predictably, the agencies might feel that such a change should be resisted. This judgment would be premature, however; the gains as well as the losses need to be considered.

The allocative efficiency standard proposed has two significant advantages. First, it makes it possible to proceed in a more self-assured way, demanding adequate relief where this is indicated. The uncertainty that characterizes present proceedings tends often to result, as it should, in relatively weak decrees. Second, even if fewer cases are brought (and this is by no means inevitable), those that are should be better cases. Filing (and winning) however many bad cases does not make for effective antitrust enforcement. It is the burden of the present argument that, if power is interpreted not in terms of latitude but rather in terms of the effective realization of a meaningful goal, an allocative efficiency standard (broadly stated and progressively refined) will increase both the power and economic importance of the antitrust enforcement efforts.

REFERENCES

1. W. J. BAUMOL, "Calculation of Optimal Product and Retailer Characteristics: The Abstract Product Approach," *J.P.E.*, Oct., 1967, pp. 674–85.
2. G. S. BECKER, "A Theory of the Allocation of Time," *Econ. J.*, Sept., 1965, pp. 493–17.
3. NORMAN BERG, "Strategic Planning in Conglomerate Companies," *Harvard Bus. Rev.*, May-June, 1965, pp. 79–92.
4. J. M. BLAIR, "The Conglomerate Merger in Economics and Law," *Georgetown Law J.*, Summer, 1958, pp. 672–700.
5. D. C. BOK, "Section 7 of the Clayton Act and the Merging of Law and Economics," *Harvard Law Rev.*, Dec., 1960, pp. 226–355.
6. R. H. BORK AND W. S. BOWMAN, "The Crisis in Antitrust," *Columbia Law Rev.*, Mar., 1965, pp. 363–76.
7. R. H. BORK, "Contrasts in Antitrust Theory: I," *Columbia Law Rev.*, Mar., 1965, pp. 401–16.
8. ———, "The Rule of Reason and the Per Se Concept: Price Fixing and Market Division, I," *Yale Law J.*, Apr., 1965, pp. 775–847.
9. ———, "The Rule of Reason and the Per Se Concept: Price Fixing and Market Division, II," *Yale Law J.*, Jan., 1966, pp. 375–475.
10. ———, "Legislative Intent and the Policy of the Sherman Act," *J. of Law and Econ.*, Oct., 1966, pp. 7–48.
11. ———, "The Goals of Antitrust Policy," *A.E.R.*, May, 1967, pp. 242–53.
12. ———, "The Supreme Court versus Corporate Efficiency," *Fortune*, Aug., 1967, pp. 92–93 *et seq.*
13. *Brown Shoe* v. *United States*, 370 U. S. 294 (1962).
14. E. H. CHAMBERLIN, "Product Heterogeneity and Public Policy," *A.E.R.*, May, 1950, pp. 85–92.
15. *Congressional Record, 21* (1890).
16. DONALD DEWEY, "Romance and Realism in Antitrust Policy," *J.P.E.*, Apr., 1955, pp. 93–102.
17. J. DIRLAM AND A. KAHN, *Fair Competition* (Ithaca, N.Y., 1954).
18. C. D. EDWARDS, statement in *Economic Concentration, Part 1: Overall and Conglomerate Aspects*, Hearings Before the Subcommittee on Antitrust and Monopoly. 82nd Congress, Washington, 1964, pp. 36–47.
19. *Federal Trade Commission* v. *Procter & Gamble Co.*, 87 S Ct. 1224 (1967).
20. FEDERAL TRADE COMMISSION, Brief Filed with the Supreme Court of the United States, *Federal Trade Commission, Petitioner* v. *The Procter & Gamble Company*, October Term, 1966.
21. ———, "Enforcement Policy with Respect to Mergers in the Food Distribution Industries," Jan. 3, 1967.
22. *Foremost Dairies, Inc.* 60 F.T.C. 944 (1962).
23. J. DE V. GRAAFF, *Theoretical Welfare Economics* (Cambridge, England, 1963).
24. H. G. JOHNSON, "The Economic Approach to Social Questions," *Public Interest*, Summer, 1968, pp. 68–79.
25. C. KAYSEN, "The Present War on Bigness: I," in the Impact of Antitrust on Economic Growth, Fourth National Industrial Conference Board Conference on Antitrust in an Expanding Economy, New York, 1965, pp. 31–38.
26. ———, "Models and Decision-Makers: Economists and the Policy Process,"

Public Interest, Summer, 1968, pp. 80–95.

27. F. H. KNIGHT, *Risk, Uncertainty and Profit* (Harper Torchbook Edition, 1965).
28. K. J. LANCASTER, "A New Approach to Consumer Theory," *J.P.E.,* Apr., 1966, pp. 132–57.
29. E. MANSFIELD, *Industrial Research and Technological Innovation* (New York, 1968).
30. E. S. MASON, forward to *Antitrust Policy,* by C. Kaysen and D. F. Turner (Cambridge, Mass., 1959).
31. F. I. MICHELMAN, "Property, Utility and Fairness: Comments on the Ethical Foundations of 'Just Compensation' Law," *Harvard Law Rev.,* Apr., 1967, pp. 1165–1258.
32. D. F. TURNER, "Conglomerate Mergers and Section 7 of the Clayton Act," *Harvard Law Rev.,* May, 1965, pp. 1313–95.
33. U. S. DEPARTMENT OF JUSTICE, *Merger Guidelines,* May 30, 1968.
34. *United States* v. *El Paso Natural Gas Co.* 376, U. S. (1964).
35. O. E. WILLIAMSON, "Hierarchical Control and Optimum Firm Size," *J.P.E.,* Apr., 1967, pp. 123–38.
36. ———, "Economies as an Antitrust Defense: The Welfare Trade-Offs," *A.E.R.,* Mar., 1968, pp. 18–36.
37. ———, "Economies as an Antitrust Defense: Correction and Reply," *A.E.R.,* Dec., 1968.
38. ———, "Administrative Decision-Making and Pricing: Externality and Compensation Analysis Applied," in *Economics of Public Output* (forthcoming).
39. ———, "Managerial Discretion and the Multi-Division Form," unpublished manuscript, Dec., 1968.
40. ———, *Corporate Control and Business Behavior* (Englewood Cliffs, N.J. forthcoming).

DISCUSSION

RICHARD B. HEFLEBOWER: The papers at this session are concerned to varying degrees with whether private efficiency is also social efficiency. Even when the survivor test is applied to properly demarcated markets, a difficult task as Bain points out, the findings indicate only the ability to survive in markets as they are. This point has been made repetitively by Stigler but there is one possible exception. If a test shows private inefficiency of the very large firm, one could infer that its scale is also economically inefficient. Comanor and Wilson's further report on the relation between profit rates and advertising expenditure suggests the uneconomic use of that revenue increasing practice because of higher scale economies in advertising than in production. Obviously the proposal of Williamson to give weight to predicted allocative efficiency of private actions that are suspect under present antitrust policy, particularly that on mergers, must include methods of distinguishing private from economic gain. But this is only part of the unresolved difficulties with Williamson's alarming proposal.

Much of Bain's paper is concerned with the heterogeneity within geographic areas and among plants and firms, to which I add three comments. Small plants located away from manufacturing centers often have lower costs than their relative scale would indicate. Wages are lower, at times a higher quality of labor is obtained, and union rules, particularly on jurisdictional issues, are less restrictive at least in application. Land is cheap, taxes low, automobile traffic is light and living costs less. Lower costs arising from these conditions represent economic efficiency, not exploitation.

More importantly, these plants often are in market segments where scale economies are not high. These include making small parts for machinery or consumer goods to move through the large volume, low price market segment. More of the part played by suppliers of the latter channel is seen, however, when they are identified as the first vertical step in a market process for which the most important firms are the mass retailers. For a variety of reasons they often find small manufacturing firms to be the best sources, whether on term contracts or in semi-bid purchasing.

These small firms often have two advantages that affect their costs favorably. They are price takers and profit maximizers in the very short calendar period. Consequently their average operating rate can approximate the low point on their average cost curve while the large rivals' average rate is well short of the comparable point on their probably lower cost curves. Second, the difference in the level of cost curves often is much less than one would infer from engineers' estimates of scale economies. For much of industry economists give undue weight to the cost effect of size of plant. Often the stability of the operating rate —or at least its predictability for weeks or months—is the greatest source of low costs. The customers of small plants, whether assemblers of finished goods or the mass retailers of consumer goods, often provide suppliers with such favorable operating conditions.

On another tack, the frequent statement that all economies of scale in form changing operations are reached at the plant level does not hold in at least two circumstances. One is the specialization of one or so of the several plants of a large company on the small volume items in the product line. The Akron plants tend to produce the specialty tires for nationwide distribution as well as the standard large volume and sizes and types to serve their region. The regional plants usually make only the latter. Often markets do not work well enough to permit a small firm to specialize on part of the line. Second, when demand slackens the multiplant firm can take much or all of the cutback in one or two of its several plants, as petroleum refiners do. Of course, such production economies must at least offset transport costs.

Finally, when testing the survival of firms, two problems arise with respect to the multimarket large enterprise, whether it operates in several geographic or product markets. Is it size in the market or overall size that is significant? If the former, a quite misleading conclusion may be reached, and it will not rest on the much debated issue as to whether a conglomerate can perform differently in one of the several markets than can single market enterprises. Instead, a conglomerate's rate of growth in one market often reflects alternatives available in other markets. For example, Stigler reports correctly that the largest petroleum refiner's share of the domestic market has fallen but no one would assert that its percentage of the world output has declined. Obviously this giant enterprise's record and the declining share of the large meat packers are examples of choosing where to invest added capital, a point on which Gort's study of directions of diversification adds convincing evidence. Because giant enterprises are conglomerates, it is not possible to use a market share test of even the private efficiency of their overall size.

Second, at the upper part of the size range the number of companies is very small and may be one only. Then it is necessary to know whether, as in another one of Stigler's cases, the largest firm's position had fallen from dominant leadership to retaining some influence in resisting price declines but without the power to discipline. In the process, it had become somnolent.

Comanor and Wilson report their latest work on the effects of advertising on profit rates and, inferentially, on the extent of monopoly. Advertising outlay is far from a satisfactory measure of the expenditures for nonprice competition. Even all selling costs would not be complete. In important commodity fields the most expensive and effective sales device is product variation, which requires added R and D and production costs. Second, it is not possible, as noted earlier, to distinguish the contribution of nonprice competition to economic efficiency compared to economic waste and consumer loss. Concern about this issue lessens where entry is easy.

This leads to the redebated question as to whether nonprice competition may establish a substantial barrier to entry. A decade ago Bain's answer was strongly in the affirmative, a view widely accepted until recently. Now the counterattack is on, epitomized by the empirical conclusions from a deductive system in McGee's rigorous criticism of a paper by Comanor at these meetings two years ago. McGee's basic point was that if product differentiation is a barrier, the reason must lie in the imperfections of the capital markets. That often used "escape hatch" when a theoretical analysis faces the empirical world involves a vague concept of a perfect capital market and of its significance for understanding the behavior of firms. Even if capital markets are found to be imperfect, that fact is a datum for the study of conditions of entry into a product market.

Whether rates of return on reported net assets, or equity, are used to indicate the height of the entry barrier or the private economies of scale, a basic deficiency of this measure must be explored. Nonprice competition outlays are composites of varying and unknown proportions of expenditures to affect the revenue of the current period, of net investment to enhance the market position so as to increase future revenues, and of gross investment to offset depreciation of the position held at the beginning of the accounting period. The authors are aware of this problem. In passing, it should be noted that the comment applies also to a substantial part of outlays for research and development and to increase crude oil producing capacity. In a period of increasing net investment in nonprice competition, charged to current expense, profits will fall. If later the net investment rate declines, or ceases, profits will rise to the extent that past investment increases gross revenues relative to current costs. But the net assets, or equity, of the firm will never include the net investment that had been charged to current expense. For that reason alone the profit rate reported in the firm's accounts will tend to be above the rate of return in highly competitive industries. If the rate proves not to be higher, the investment out of what could have been larger reported earnings in the past was a mistake.

Williamson's paper may be in response to the lawyers' frequent request, taunt might be the better word, directed to economists to develop meaningful economic standards for antitrust. Actually, the author desires to develop policy to accord with economic theory. Hence the emphasis on allocative efficiency, a policy guide traced to Bork's paper of three years back. But Williamson goes beyond the usual economic concept of allocative efficiency to incorporate sociopolitical elements. One mentioned is public discontent, hardly a quantifiable variable, by means of which what has been termed "populism" would become economically respectable. The author might well add public concern about the syndicalist trend in our society if his proposal for a predicted allocative efficiency test of mergers were adopted.

The fundamental character of the change in antitrust policy proposed by Williamson should be underscored. Instead of being aimed primarily to maintaining a system of satisfactorily competitive markets, from which economic allocative efficiency is expected to emerge, he proposes a case by case estimate of the trade-off between effects on competition and the predicted, and I emphasize that word, effects on allocative efficiency. This is the very kind of decision making, that of a trade-off, which Bork argued the courts are not qualified to make and studiously avoid.

Williamson's proposal is directed primarily to the current policy on mergers, the phase of antitrust policy for which his suggestion is least needed. The burden of proof is on those that argue that mergers improve allocative efficiency; so far reliable supporting evidence is lacking in the cases against which the action has been initiated. In contrast it has not been shown clearly in these cases that the acquiring firm could not more slowly, but probably more certainly, achieve the economic efficiency results by internal expansion.

Surely an even better case for Williamson's proposal could be made for applying his rule to many cases of collusion, as have the British. Bitter price competition in a high capital-output ratio industry forms a poor basis for employing the most efficiency production functions and for progressiveness.

To an increasing degree in recent years the alleged social benefits have been judged by the British Court to outweigh the monopoly effects, with the result according to Caves in the recent volume, *Britain's Economic Prospects,* that economic growth has been retarded.

Economists could not help much in applying Williamson's proposed policy to particular cases. They do not have definitive theory, and certainly not that which has been tested, for important components of the criteria involved in the trade-off. Nor could economists go far in quantifying the components of each criterion let alone provide the weights each of the variables should have in the composite standard. The difficulty is that the future effects must be estimated.

Williamson recognizes all of this but is optimistic about what can be done to approach quantification of an economic second-best judgment by following four steps to be taken over time. Presumably, success in each step would provide confidence in the map to that point, and indicate good and bad roads in the next area. In contrast I suggest that a clear start on the four-stage route would make stopping or even reversing the course politically and administratively impossible. The best policy is still to rely on one basic criterion, of the effect on competition even though, when viewed against certain standards, it appears as a policy second best.

LESTER G. TELSER: Because my sympathies are always with those who do empirical work I shall confine most of my remarks to Comanor and Wilson's paper. Their central proposition, as I understand it, postulates a positive association between advertising and the size of firms. In a given product class the bigger firms are the ones that advertise more relative to their sales. This is not an implausible proposition in some respects though it is complicated if there are increasing returns to advertising. Its plausibility derives from the fact that advertising is a sales technique best suited for those products with a mass market. For example, using cigarette data[1] I found a positive relation between market share and relative advertising outlays. A more pertinent example is beer. Nationally advertised brands of beer generally have larger sales and higher ratios of advertising to sales than locally advertised brands. More generally, this proposition implies that within a product class, the nature of the advertising by brand differs so that the less intensively advertised brands rely more on local media than the more intensively advertised brands.

---

[1] L. G. Telser, "Advertising and Cigarettes," *J.P.E.*, 1962, pp. 471–99.

There are some other complications because advertising intensity also depends on product newness. Thus new products gain acceptance by advertising more relative to sales than the more established products. This factor, product newness, needs also to be taken into account when studying the relation between advertising intensity and relative size. Thus to avoid such complications one may prefer to study only the well-established brands within a product class.

However, none of these routes is the one taken by Comanor and Wilson. They broaden the hypothesis by asserting that advertising not only affects quantity sold but also it affects price. Thus they postulate that advertising permits a given quantity to be sold at a higher price in such a way as to increase the rate of profit. As a result they are led to measure how advertising affects the size of firms through a relation between size and the profit rate. At this point I part company from them. Their analysis fails to convince me for several reasons that I shall now explain.

The authors determine the "minimum efficient size" from the profit rate by estimating a regression equation (1) which makes the profit rate a function of $1/S$ and $1/S^2$. This regression is estimated for forty-one consumer product industries, the same ones I used in my paper, "Advertising and Competition."[3] Each of the forty-one regressions is based on about eight degrees of freedom, the eight size classes by industry including the open ended class over $250 million dollars. Table 1 shows that the standard error of the estimate of the minimum efficient size exceeds the estimate itself in nine cases out of twenty-nine. In the remaining twelve cases the profit rate approaches the postulated asymptote from above instead of from below as their hypothesis requires. Thus about half the industries violate the authors' premise on their own evidence. Moreover in six cases the estimated minimum efficient plant size actually exceeds the estimated minimum efficient firm size. Using these precarious estimates the authors explain the ratio of minimum efficient firm size to market as a function of two variables, advertising intensity measured by the ratio of advertising to sales and the ratio of minimum efficient plant size to market. This regression has a statistical flaw known as spurious correlation because one of the explanatory variables, if not both, has the same denominator as the dependent variable. Therefore, even if there were no relation between the variables, measurement error alone could give positive regression coefficients. This point and the fact that in any case the coefficients are small relative to their standard errors prevents me from taking seriously their results.

The other main piece of evidence tries to explain the difference in the profit rate between the larger and the smaller firms in a product class using three explanatory variables of which one is the advertising to sales ratio.

To these results my objections rely on the underlying economic theory. The problem here is whether or not the measured profit rate is a biased estimate of the true profit rate. The measured profit rate is the ordinary accounting measure of profit after taxes divided by stockholder equity. The latter is total assets less debt. The problem is that stockholder equity understates the true capital of the firm and that the measured profit understates the true profit. Both are the effects of the same cause; namely, the tendency to omit intangible capital. Stockholder equity includes only tangible capital and the profit is too low because in effect it allows a 100 percent rate of depreciation of intangible capital. The point is well illustrated by advertising. In my article, "Advertising and Cigarettes," I found that the effects of advertising last for a period of time after the actual outlays so that advertising builds up a form of capital. The relation between the true and the measured rate of return is best expressed in a simple formula.

---

[3] L. G. Telser, "Advertising and Competition," *J.P.E.*, 1964, pp. 537–62.

Let A = advertising capital and d_A its rate of depreciation
K = tangible capital and d_K its rate of depreciation

The true rate of return denoted by r is defined as follows:

$$(1) \quad r = \frac{\text{Sales} - \text{Variable cost} - d_A A - d_K K}{A + K}$$

and the measured rate of return on tangible capital is defined by

$$(2) \quad \hat{r} = \frac{\text{Sales} - \text{Variable costs} - a - d_K K}{K}$$

where a is the advertising outlay. Notice that in (2) the entire advertising outlay is subtracted in obtaining the profit after taxes while in (1) only the depreciation of the advertising capital is subtracted from sales. It readily follows from (1) and (2) that

$$(3) \qquad r - \hat{r} = (\dot{A}/A - \hat{r})w_A$$

where

$$w_A = A/(A + K), \qquad A = \frac{dA}{dt},$$
$$\text{and } A - a = AdA.$$

Formula (3) is most instructive. It shows that the true rate of return is above the measured rate of return if and only if the rate of growth of advertising capital is above the measured rate of return. In particular the measured rate is always above the true rate in the steady state when the advertising capital is constant.

A positive coefficient for the advertising to sales ratio in the Comanor-Wilson regression may indicate that the measured profit rate is an upward biased estimate of the true profit rate where the bias varies directly with the ratio of advertising outlays to sales. In other words the advertising to sales ratio may serve as a proxy for the missing intangible capital in the form of advertising. In particular the proportion of missing capital may be higher in the larger firms for which there is a higher ratio of advertising to sales. In any event without evidence to indicate how the measured profit rate is related to the true profit rate, there is no reason to accept the authors' conclusion that advertising raises the firm's profit rate. All that we may conclude is that we have not measured the profit rate correctly because we have omitted intangible capital.

That conventional accounting measures of capital are too low is shown by the fact that the market value of firms given by the value of their outstanding stock usually exceeds the book value of their total assets. In addition according to *Fortune* the ten most profitable firms ranked according to the rate of profit on stockholder equity generally includes five or so that either advertise heavily or spend lavishly on reasearch and development. Since R and D outlays are also a form of intangible capital, I would cite this as additional evidence that something is wrong with our measures of capital and with measured profit rates and not that advertising and R and D outlays are a means to especially high profits. However, if with better measures of capital it should turn out that rates of return on such intangible capital are higher than on tangible capital at the margin, then we would face the problem of explaining why this is so. Are these investments subject to greater risk? But first the evidence.

I have only one comment for Professor Bain. Since he addresses most of his paper to a criticism of Professor Stigler's classic article on the nature of the returns to scale, it is a pity that he does not provide the reader (or listener) with the most recent reference; namely, Stigler's new book, *The Organization of Industry*,[3] wherein are Stigler's own comments on the subject. In particular Stigler points out that the persistence of firms in certain size classes does not tell us about the turnover within that class so that we cannot infer whether or not these firms pass the survivor test. I missed this most important point in Bain's discussion. I can hardly imagine anyone nowadays giving estimates of the nature of the returns to scale without paying at least lip service to the ingenious test of survivorship due to Professor Stigler.

---

[3] George J. Stigler, *The Organization of Industry* (Irwin, 1968), esp. pp. 71–94.

WAGE-PRICE DYNAMICS, INFLATION, AND UNEMPLOYMENT

THE PHILLIPS CURVE AND THE DISTRIBUTION OF UNEMPLOYMENT*

By G. C. ARCHIBALD
University of Essex

Several economists, including the present writer,[1] have suggested in recent years that a reduction in the dispersion of unemployment (specifically, the geographical dispersion in the U.K.) would, *ceteris paribus*, reduce the rate of wage change associated with values of the national unemployment percentage. The relationship between the dispersion of unemployment and the rate of wage change is interesting in its own right; it is of considerable importance to policy; and it affords a test of part of Lipsey's theory of the Phillips curve. In the present paper we shall, first, briefly review the relevant theory. We shall then consider some empirical results for the U.K., and for the U.S. (postwar). For the U.K., we use dispersions on a geographical and an industrial basis; for the U.S. on geographical basis only. The empirical evidence is not, alas, altogether clear, but some tentative conclusions are suggested. These may be briefly summarized.

1. The existence of Phillips curves, in the sense of well-established relations between the rate of change of wages and the level of unemployment, is fairly clear for the postwar period.
2. The first and second moments of the distribution of unemployment in the samples considered are fairly well correlated but poorly correlated with the third moment (the correlations vary from sample to sample in a fashion which will be discussed below).
3. There is therefore some problem of multicollinearity, but it seems clear that the second moment of the distribution of unemployment does have a positive effect on the rate of change of wages, in the postwar samples at least.[2]
4. The moments are generally not correlated with the rate of change of the mean (U-dot) in the samples considered, contrary to Lipsey's auxiliary hypothesis that U-dot served as a proxy for dispersion in the fitted equations.
5. U-dot in any case appears to be unimportant for the U.K. data samples considered here. The evidence for the U.S. is less clear.
6. For the U.K. the simple correlation between the rates of wage change and of price change is fairly high, much higher than for our short U.S. postwar sample period. If the former depends upon excess demand in the labor market and the latter on excess demand in the goods market and excess demands for labor and goods are highly correlated, as we should suppose, this is unsurprising. The question, however, is whether we can obtain more information about the rate of change of

---

* This paper in part reports on some work done during my tenure of a Ford Foundation Faculty Fellowship at Northwestern University, 1967-68. I take this opportunity to express my great indebtedness both to the Foundation and to my host University. Neither, of course, bears any responsibility for what is said here. The work on the U.K. data also reported here was made possible by a grant from the S.S.R.C., to which I am similarly indebted. Mr. J. W. Perkins has been an admirable research assistant. I have had both advice and practical help at every stage from Professor F. P. R. Brechling on such a scale that, were he not a discussant, I should try to enlist him as coauthor. Neither he nor Professors B. A. Corry and R. G. Lipsey, to whom I am also indebted, bear any responsibility for errors.

[1] *Oxford Economic Papers*, Mar., 1967.

[2] The interwar period in the U.K. is currently being investigated and will be reported on separately. Preliminary results are extremely puzzling.

wages from variables specific to the labor market. For the postwar U.K. samples we clearly can. For the short postwar U.S. samples we can, too, but the rate of change of prices is anyhow less important in the fitted wage equations.

7. The positive correlation between the first and second moments of unemployment suggests an association between "structural" and "demand-deficient" unemployment which might induce skepticism about a debate on the issue "structural versus demand-deficient" explanations of unemployment; but we shall not pursue this topic here.

One further introductory remark may be in order. There has recently been considerable debate[3] over the question of whether the Phillips curve presents a stable frontier to the policy-maker. There is no inconsistency between the proposition that the curve may be shifted by changing expectations and the proposition investigated here, that it may be shifted by changing the dispersion of unemployment.

I

1. Some of the explanatory variables appearing in fitted wage-change equations, including Lipsey's own, are prima facie inconsistent with the simple excess demand model. It was to deal with one of these "intruders," U-dot, that Lipsey proposed the dispersion hypothesis.[4] (Space does not permit consideration of the other intruders here.)

The Phillips-Lipsey hypothesis is that $\dot{w}/w$ is a function (perhaps linear) of excess demand, X, and that there exists a stable (but presumably nonlinear) transformation between excess demand and unemployment. It follows from elementary price theory that, if $\dot{w}/w=f(X)$, we should have $f(O)=0$. Thus we should not find any variable, such as prices or productivity, which affects either the demand curve or the supply curve, or both, apparently exercising an independent effect on $\dot{w}/w$: X should be the only argument of the function, the slope of which may, however, be affected by extraneous variables, as Lipsey pointed out. (He also pointed out that, since we do not know the excess demand-unemployment transformation, we cannot detect the slope-changing effect of, e.g., trade-union strength.)

This simple excess demand hypothesis, although conventional and common in the literature, may be criticized as *ad hoc* since it is not derived from any theory of the behavior of the market agents in disequilibrium situations. Recent work by Phelps (*op. cit.*) and important unpublished work by Holt and Mortensen are directed to the derivation of a dynamic theory of maximizing behavior under disequilibrium and, in the last two cases, under uncertainty as well. Phelps's work, in particular, provides an alternative to Lipsey's explanation of U-dot: that it is a joint proxy, with U, for X itself. Here we follow Lipsey's hypothesis, particularly because, while we do not find U-dot significant in the fitted equations, we do find the dispersion of unemployment (for which Lipsey suggested that U-dot was a proxy) exercising a significant effect.

2. To account for the "loops" round his

---

[3] For the notion of a policy frontier, see particularly Phillips, *Economica*, 1958, Lipsey, *Economica*, 1960, and in *Employment Policy and the Labor Market*, ed. A. M. Ross (1965), Perry, *Rev. of Econ. Studies*, 1964, and Brechling, *J.P.E.*, 1968. For the expectational hypothesis, see particularly Phelps, *Economica*, 1967, and *J.P.E.*, 1968, and Friedman, *A.E.R.*, 1968.

[4] The existing empirical results give a formidable list of intruders. Besides U-dot (Lipsey) and the rate of change of prices, which appears in practically every reported equation, we have profit and its rate of change—in the U.S. (Perry, *op. cit.*) but not in the U.K. (Lipsey and Steuer, *Economica*, May, 1961), trade-union strength in the U.K. (Hines, *Rev. of Econ. Studies*, Oct., 1964, and *Rev. of Econ. and Statis.*, Feb., 1968) and now in the U.S. (Gale Pierson. *A.E.R.*, June, 1968), and productivity in the U.S. (Kuh, *Rev. of Econ. Studies*, Oct., 1967).

curve, Phillips (*Economica*, Nov., 1958) originally suggested that, when excess demand was increasing (unemployment falling), employers would anticipate tightening labor markets, and hence wish to increase their work forces while they could, whereas in a downswing, with easing markets, they would expect yet greater ease and tend to postpone recruitment. Lipsey argued that, if they altered their hiring, they must directly alter excess demand, so that the appearance of its rate of change in the fitted equation was still not consistent with the adjustment hypothesis. He therefore produced his familiar aggregation hypothesis, which I shall briefly summarise.

Assume (1) that micromarkets have their own identical micro-Phillips curves and (2) that they too are convex. Then (3) if unemployment is not the same in the micromarkets, the fitted macrorelationship, a weighted average of individual-market values of $\dot{w}/w$, will be above the individual market relationships. Further (4) observed $\dot{w}/w$ will be higher for given U when the dispersion of unemployment is high than when it is low. Hence (5) anticlockwise loops, apparently explained by U-dot, are consistent with the excess-demand hypothesis if dispersion is positively correlated with U-dot, so that the latter has served as proxy for the former. Thus Lipsey's attempt to "save" the original hypothesis has the admirable feature of leading to a further testable hypothesis.

Points (4) and (5) require elaboration: on what conditions do we obtain the prediction that a *ceteris paribus* reduction in dispersion will reduce $\dot{w}/w$? Lipsey assumed for convenience of exposition that the microcurves were identical as well as convex. We may easily see that the first assumption is not necessary, and that strict convexity is neither necessary nor sufficient.

Let U, U_A, and U_B denote percentage unemployment in the whole economy, in region A and in region B, respectively, and let these two regions (or markets) comprise the whole economy, so that we may write

$$(1)\qquad U = \alpha U_A + (1-\alpha)U_B \qquad 0 < \alpha < 1$$

where α is the fraction of the whole labor force in region A. If we have two convex Phillips curves, we may also write

$$(2)\qquad \frac{\dot{w}}{w} = \alpha(a_1 U_A^{-b_1} + C_1) + (1-\alpha)(a_2 U_B^{-b_2} + C_2)$$

where $\dot{w}/w$ is the national rate of change of wages. Now, a *ceteris paribus* reduction in dispersion means that we reduce unemployment in the high unemployment region, say region A, increasing it in B by the amount required to keep U constant. So we write

$$(3)\qquad dU = \alpha dU_A + (1-\alpha)dU_B = 0$$

Also from (2)

$$(4)\qquad d\left(\frac{\dot{w}}{w}\right) = -\alpha b_1 a_1 U_A^{-(1+b_1)} dU_A - (1-\alpha) b_2 a_2 U_B^{-(1+b_2)} dU_B$$

Using (3)

$$(5)\qquad \frac{d\left(\dfrac{\dot{w}}{w}\right)}{dU_B} = -\alpha b_1 a_1 U_A^{-(1+b_1)}\left(1 - \frac{1}{\alpha}\right) - (1-\alpha) b_2 a_2 U_B^{-(1+b_2)}$$

and, rearranging,

$$(6)\qquad \frac{d\left(\dfrac{\dot{w}}{w}\right)}{dU_B} = (\alpha - 1)\left\{b_2 a_2 U_B^{-(1+b_2)} - b_1 a_1 U_A^{-(1+b_1)}\right\}$$

For (6) to be negative, as required, the condition is that the term in braces be positive. But comparison with (2) shows that this condition is that the Phillips curve be steeper in region B than in region A, given that unemployment in B is lower than in A. We also see from (6) that the direction of the dispersion effect, though not, of course, its magnitude, is independent of the weights. To express the result directly in terms of dispersion, we merely note that

$$\frac{d\left(\frac{\dot{w}}{w}\right)}{d\sigma^2} = \frac{d\left(\frac{\dot{w}}{w}\right)}{dU_B} \cdot \frac{dU_B}{d\sigma^2}$$

and, on the assumptions that $U_A > U_B$, $dU_B > 0$, and $dU = 0$, obviously $(d\sigma^2/dU_B) < 0$. As for the condition on the slopes, it would be satisfied in the case of parallel displacement. In the special but interesting case in which $b_1 = b_2 = 1$ it requires $a_1 = a_2$. The slope conditions may appear reasonable; at all events they are less restrictive than the requirement of identity.

The case of linearity is given added interest by the finding (see Brechling *op. cit.*) that a linear macrocurve fits better in the U.S. than a nonlinear one. This case is now easily disposed of. We have

$$\frac{\dot{w}}{w} = \alpha(d_1 U_A + C_1) + (1 - \alpha)(d_2 U_B + C_2) \tag{7}$$

where

$$0 < \alpha < 1 \quad \text{as before}$$

but

$$d_1 < 0, \quad d_2 < 0$$

Repetition of the argument that led to (6) gives

$$\frac{d\left(\frac{\dot{w}}{w}\right)}{dU_B} = (\alpha - 1)(d_1 - d_2) \tag{8}$$

(8) is negative if and only if $d_1 - d_2 > 0$; i.e., if the high unemployment region A has a flatter Phillips curve at least in the relevant range than the low unemployment region. But this is just the condition in the strict convexity case. In the case of identical linear curves, of course, there is no dispersion effect.

It remains to check which way the loops might go in the linear case. Clearly, if the dispersion effect is positive, which means that markets in which unemployment tends to be higher have relatively flatter Phillips curves, then we require, as in the convexity case, Lipsey's hypothesis that dispersion increases when U falls if we are to account for anticlockwise loops.

3. We now consider possible ways of testing Lipsey's hypothesis. One way is clearly to look for individual-market Phillips curves. The obstacle is that the available data are usually for geographical or industrial subaggregates that need by no means coincide with "true" labor markets. The approach taken here is therefore to look directly for the dispersion effects discussed in 2 above. Thus we fit time series equations in the manner of Lipsey, Perry, etc., but use the dispersion of unemployment at each time as an additional explanatory variable. The measure of dispersion used is simply the variance of unemployment computed from whatever data are available, whether geographical or industrial. This leads to multicollinearity, which might be avoided by a different choice of variables; but see Section IV below. We thus assume, without test, that although the basis of reporting may not coincide with individual markets, the dispersion computed from available data will serve as a good proxy for the dispersion of excess demand over the true markets. In general, we find that the first and second moments of the distributions are positively correlated (in this context, the national unemployment percentage, macro-U, is the first moment of the distribution). The

transformation between excess demand and measured unemployment, however, presents us with a further complication: at low values of unemployment, the dispersion of unemployment is likely to understate the dispersion of excess demand, since excess demand can have any real value whereas unemployment cannot fall below zero. Thus suppose that the distribution of excess demand were unskewed. The distribution of measured unemployment would necessarily be skewed at low levels of mean unemployment and could only be unskewed at high unemployment levels. Thus it appears that the transformation alone would cause the observed distribution to display a skewness negatively associated with its own mean; and the distribution of excess demand might itself not be symmetrical. In an attempt to allow for this effect, the third moments have been calculated in each case, and a measure of skewness used in the multiple regression equations. For the U.K. data, it proves that the second and third moments are positively correlated. This seems contrary to expectations. The measure of skewness is not, in fact, significant in most of the equations (but where it is, it has a negative sign). All this is a little confusing. It is possible that the relationship between the measures of dispersion used here and the dispersion of true excess demand is such that we have a case of errors in variables; but this possibility has not been explored.

II

1. We now report our results for the U.K., 1950–66, an annual model on a regional basis. Detailed sources are given in the notes to Table 2, which reports the more important regression results. The correlation matrix is given in Table 1. U is here the mean, σ^2 the regional variance each year, i.e., $\sigma^2=\sum f_i(U_i-\bar{U})^2$ where f_i is the proportion of the labor force in the ith region and U_i its unemployment percentage. γ is a measure of skewness ($m_3/m_2^{3/2}$: see Kendall and Stuart, *The Advanced Theory of Statistics*, Vol. I, 2nd ed., p. 85). All rate-of-change variables are first central differences. A number of experiments were made that need not be reported in detail. First, the evidence is clearly in favor of nonlinearity in U: all equations were rerun with U in place of U^{-1}, with worse results. Second, the evidence is consistent with some sort of Phillips curve: equations including P-dot, but excluding U^{-1}, were worse. Third, a time trend proved to be insignificant. Fourth, there are several data breaks, of which the most serious was caused by a boundary shift in 1964. The shifts do not affect measured U, but might affect measured σ^2. Consolidation to nine regions avoided all but the 1964 shift. A dummy variable for that shift proved not to be significant, but, as may be seen from equation (3), leads to a slight increase in the significance level of other coefficients.

The results appear to be fairly clear and encouraging. The first two moments of the distribution are well correlated, but not so well correlated with γ (hardly surprising: there are only 9 observations each year from which to compute the third moment). The correlation between wage change and price change is high, and between wage change and σ^2 trivial. Contrary to Lipsey's hypothesis, σ^2 and U-dot are uncorrelated. In equation (1), a Phillips-Lipsey equation,[5] the only significant variable is P/P! Replacing $\dot{U}/U$ with σ^2, equation (2) gives the predicted result: σ^2 has a significant positive coefficient, as well as U^{-1} and $\dot{P}/P$. Equations (3), (4), (5) and (6) show that neither $\dot{U}/U$ nor γ is significant in any combination of variables, but that

---

[5] Direct comparison with Lipsey's results is difficult, since he pooled parts of the interwar and postwar periods. It seems, however, that his coefficient for U^{-1} was scarcely significant, although one for U^{-4} was significant, and that $\dot{U}/U$ was not in fact significant.

TABLE 1

U.K. POSTWAR REGIONAL MODEL: CORRELATION MATRIX

	σ^2	$\dot{U}/U$	γ	U^{-1}	$\dot{P}/P$
$\dot{U}/U$	−.095				
γ	−.307	.309			
U^{-1}	−.803	.249	.462		
$\dot{P}/P$	−.165	.372	−.056	.538	
$\dot{W}/W$	−.126	.206	.215	.608	.807

U^{-1}, σ^2 and $\dot{P}/P$ always are.

On this evidence, we might conclude that there is a Phillips curve, and that a reduction in the regional dispersion of unemployment in the U.K. would move the fiscal policy frontier in a favorable direction.[6]

It will be noticed that no attempt has been made to allow for periods of wage freeze. This topic is beyond the scope of

TABLE 2

U.K. POSTWAR REGIONAL MODEL: REGRESSION EQUATIONS

(T-values beneath coefficients; $n=17$)

Equation Number	Const.	U^{-1}	$\dot{P}/P$	$\dot{U}/U$	σ^2	γ	D	$\bar{R}^2$	DW
(1)	1.118 (1.2)	2.185 (1.4)	0.555 (3.9)	−0.007 (0.8)				.64	1.16
(2)	−4.398 (2.3)	8.321 (3.6)	0.315 (2.5)		1.709 (3.1)			.78	1.19
(3)	−4.501 (2.5)	8.774 (4.0)	0.294 (2.5)		1.683 (3.3)		−0.821 (1.7)	.81	1.48
(4)	−4.319 (2.0)	8.063 (2.6)	0.327 (2.1)		1.673 (2.6)	0.035 (0.1)		.77	1.22
(5)	−4.636 (2.4)	8.475 (3.7)	0.348 (2.7)	−0.008 (1.1)	1.729 (3.2)			.79	1.28
(6)	−4.273 (2.1)	7.084 (2.3)	0.422 (2.5)	−0.011 (1.3)	1.536 (2.4)	0.193 (0.7)		.78	1.53

Notes:

1. $\dot{P}/P$ is first central difference, where P is a twelve-month average of the Ministry of Labour's Index of Retail Prices, the weighting of which changed several times during the period. Source: *Annual Abstract of Statistics.*
2. $\dot{W}/W$ is first central difference, where W is the twelve-month average of the Index of Weekly Wage Rates. Source: *Annual Abstract of Statistics.*
3. To measure U^{-1} and $\dot{U}/U$ the national percentage unemployment rate was used. This is not consistent with the weighted regional mean used to compute σ^2, but the difference is, in fact, trivial.
4. $\sigma^2_t = \sum_i f_i(U_{it} - \bar{U}_t)^2$ where f_i is the proportion of the labor force in the i^{th} region and U_{it} its percentage unemployment at time t. Regional boundaries were altered several times during our period. To obtain consistency, some grouping was necessary. (*a*) "Eastern and Southern" is a single region from 1962. The two regions were therefore combined, 1950–61. (*b*) A consolidated region, "Midlands," is composed of "Midlands, North Midlands, East and West Ridings," 1950–61, "Midlands, Yorkshire and Lincolnshire," 1962–64, "Midlands, Yorkshire and Humberside," 1965, and "West Midlands, East Midlands, Yorkshire and Humberside," 1966–67. In 1958 Dorset (excluding Poole) was transferred from "Eastern and Southern" to "South Western." This was ignored. Alteration of the boundaries of London in 1964 produced a large reapportionment of areas between "London and South East" and "Eastern and Southern." As reported in the text a dummy variable was used to catch any shift in σ^2.

---

[6] A. P. Thirlwell has been kind enough to show me an unpublished paper in which he fits a regional model similar to that reported here, but to better data, since he obtained consistent quarterly regional data from the government, and also an industrial model similar to that reported in the next section. His results seem to be in broad agreement with those reported here. To avoid multicollinearity he uses σ/μ in place of σ^2; but see Section IV below.

TABLE 3

U.K. Postwar Industrial Model: Correlation Matrix

	σ^2	$\dot{U}/U$	γ	U^{-1}	$\dot{P}/P$
$\dot{U}/U$	.103				
γ	.700	.115			
U^{-1}	−.485	.249	−.554		
$\dot{P}/P$	.176	.372	−.204	.538	
$\dot{W}/W$	.191	.206	−.276	.608	.807

this paper. Examination of the residuals from equation (2) revealed nothing worth reporting.

2. Results for the U.K. postwar annual model using industrial instead of regional dispersion are very similar to those just given. The dispersion is calculated from industry unemployment according to the Standard Industrial Classification. The correlation matrix is given in Table 3, and the more important regression results in Table 4. Detailed sources are given in the notes to that table. Again, there are data breaks which might affect σ^2 and γ. It was possible to calculate σ^2 on the old and the new basis for 1964. It differed by only 2 percent, whence the 1964 break was ignored. The effects of including a dummy variable for the 1958 reclassification can be seen in equation (3). The same experiments reported for the regional model (dropping U^{-1}, replacing it with U, adding a time

TABLE 4

U.K. Postwar Industrial Model: Regression Equations

(T-values beneath coefficients; $n=17$)

Equation Number	Const.	U^{-1}	$\dot{P}/P$	$\dot{U}/U$	σ^2	γ	D	$\bar{R}^2$	DW
(1)	1.118 (1.2)	2.185 (1.4)	0.555 (3.9)	−0.007 (0.8)				.64	1.16
(2)	−0.155 (0.1)	4.823 (2.5)	0.340 (2.2)		0.269 (2.0)			.71	1.15
(3)	0.853 (0.8)	4.151 (2.3)	0.302 (2.2)		0.208 (1.7)		−0.739 (1.9)	.76	1.48
(4)	0.961 (0.8)	4.783 (2.8)	0.247 (1.7)		0.461 (2.9)	−0.587 (1.9)		.76	1.53
(5)	−0.450 (0.4)	5.095 (2.7)	0.370 (2.4)	−0.010 (1.1)	0.287 (2.1)			.72	1.32
(6)	0.660 (0.5)	4.953 (2.8)	0.275 (1.8)	−0.006 (0.7)	0.457 (2.8)	−0.523 (1.6)		.75	1.62

Notes:

1. $\dot{P}/P$, $\dot{W}/W$, and U as in Table 2.

2. σ^2. Numbers of unemployed according to the S.I.C. Classification are available monthly, based on the industry in which the individual last worked before becoming unemployed. June figures were used. The classification of employees by industries is based on a count of National Insurance cards exchanged each June (believed to be a random sample of one-quarter of all cards). For 1962–66 end May figures were used, for the remaining years June, in order to compute percentage unemployment by industries. The problem of "Persons not classified by industry" is discussed in the text. The S.I.C. was altered in 1958 and 1964. In the interests of consistency, industries which were separately classified before but not after, or after but not before, 1958, were consolidated with "Other Manufacturing Industries." Similarly, the categories "Professional Services" up to 1958 and "Professional and Scientific Services" thereafter were consolidated with "Miscellaneous Services," together with "Ex-service personnel not otherwise classified by industry." These consolidations, together with the treatment of the unclassified unemployed, leave us with twenty-one industries throughout the period. The results of including a dummy variable for the 1958 break are reported in the text. In 1964 data were reported on both bases. The change altered σ^2 by less than 2 percent in that year, whence the break was ignored. Source: *Annual Abstract of Statistics.*

trend) were repeated, with similar results. Table 4, in which equation (1) is identical with (1) in Table 2, tells its own story: U^{-1} is significant and σ^2 has a significant independent influence, at least when γ is included, but γ itself does not. (The correlation between U^{-1} and γ is of opposite sign in the regional and industrial models!)

There are a few special points to notice about the industrial model. The first is that inclusion of the year 1949 (not possible for the regional model) gives much worse results than those reported, presumably due to the persistence of wartime controls, the change in the school leaving age in 1947, wage restraint, etc. The second is that σ^2 is an extraordinarily large number in 1952, due to high unemployment in the textile trade in July, which is the reporting month for calculation of σ^2. The residual for 1952 is not conspicuously large. The third is that there is a serious classification problem, which does not occur in the regional model. The unemployed are allocated to the industry of previous employment. There is a large category of unemployed "Not classified by industry" —school leavers and persons unemployed for a year or more due to temporary withdrawal from the labor force. This category cannot be omitted; but where should it be allocated for the computation of σ^2? This difficulty has been arbitrarily solved by allocating the unemployed in this category over all industries according to their weights. It seems likely that it accounts for the industrial model not fitting quite as well as the regional model.

III

1. For the U.S. we have used geographical measures of dispersion in the postwar world for which we may obtain quarterly unemployment figures by states and by the 150 major labor areas. Another possible model would use the dispersion of unemployment among the ten major industry groups; but this has not been tried. The states results are reported here, the 150 areas next. Sources, etc., are given in the notes to the tables, but some explanation is required.

(*a*) As can be seen from Table 6, dummy variables were employed (second, third, and fourth quarters), and seasonality proves to be very strong. Hence Table 5 gives not the correlation matrix but the matrix of partial correlation coefficients from multiple regressions between the independent variables with the dummies. It appears that the seasonality may be as much in the independent variable, $\bar{U}$, as in the dependent variable.

(*b*) The unemployment figures are inclusive. Results are therefore not directly comparable with previous results (Perry, etc.) for manufacturing only. This is inconvenient, but cannot be avoided if a measure of dispersion is to be obtained.

(*c*) There is no appropriate index of the money wage rate. Two dependent variables were tried: the percentage change in nonmilitary employee compensation per man; and the percentage change in earnings per man week in manufacturing only. Results were broadly similar with both variables, but uniformly better, as might be expected, with the former, which is, accordingly, the only one reported here.

(*d*) Weighted average unemployment, $\bar{U}$, computed from the geographical data does not perfectly agree with the macro-percentage, nU, although the correlation is very high. In the interests of consistency with the other moments, $\bar{U}$ is preferred here.

(*e*) The requirement that we use the set of moments precludes any adjustment of nU or $\bar{U}$ for changing participation rates (see Simler and Tella, *Rev. of Econ. and Statis.*, Feb., 1968).

(*f*) The model fitted is linear in $\bar{U}_t$ and $\bar{U}_{t-1}$ (quarterly), since several writers including Brechling have reported the superiority of the linear over the nonlinear form for the U.S.

TABLE 5

U. S. STATES MODEL: MATRIX OF PARTIAL CORRELATION COEFFICIENTS

(From multiple regression with quarterly dummies)

	nU	$\overline{U}$	σ^2	M_3	γ
$\overline{U}$	.945				
σ^2	.615	.745			
M_3	−.362	−.261	.249		
γ	−.578	−.496	.049		
$\Delta\overline{U}$			.54	.194	.08

From Table 5 we have the highest correlation between σ^2 and U-dot of any sample reported. $\overline{U}$ and σ^2 are well correlated, hence also $\overline{U}$ and γ; but σ^2 and the third moment are poorly correlated.

Equation (1) in Table 6 is in fact fitted to the quarterly data from 1950-I through 1967-IV: it is presented for comparison with equation (2), which employs the same variables for the short period 1959-I to 1967-I for which we can estimate dispersion. Equations (1) and (2) appear consistent: nU_0 and nU_{-1} are significant, $\dot{P}_0$ and $\dot{P}_{-1}$ are not (although the latter becomes significant if $\dot{P}_0$ is omitted). Unfortunately, the coefficients on nU_0 and nU_{-1} are of opposite sign and nearly equal magnitude, suggesting a rate-of-change effect, and thus reintroducing the intruder we disposed of in the U.K. case. The rate-of-change effect is less obvious in the equations including σ^2 and using $\overline{U}$: a dying weights effect looks plausible, but we recall the correlation between σ^2 and $\dot{U}$. In an endeavor to check this, equations were fitted to $\overline{U}_{-2}$ and $\overline{U}_{-3}$ as well as $\overline{U}_{-1}$ and $\overline{U}_0$. The results (not reported in Table 6) are inconclusive: the coefficients die, but so also do the T-statistics.

Equations (3) through (6) in Table 6 give the more important results. There is a problem of multicollinearity, but we may conclude that σ^2_{-1} has the predicted positive effect, and lagged skewness perhaps has some negative effect. $\overline{U}_0$ is always significant: there appears to be a Phillips curve, and $\dot{P}$ is not important.

2. The areas model is disconcerting and, at first sight, disappointing. One's expectation is that it should be possible to estimate the moments better from 150 cells than for the states or the nine U.K. regions. This is the case if the observations for the first quarter of 1964 are dropped: this quarter throws the correlation matrix out badly (the second and third moments are, for some reason, extraordinarily large), as may be seen from comparison of Tables 7-A based on all observations and Table B, omitting 1964-I. The change from positive to negative in the association between σ^2 and γ is disconcerting! Table 7-B, however, suggests serious multicollinearity, and indeed it is virtually impossible to obtain a significant coefficient in the multiple regression equations. There are two ways of finding a significant coefficient. One way is to use the less appropriate dependent variable, percentage change in earnings per man week in manufacturing only. Then the lagged values of any of the three moments is significant, or nearly so, provided the other two are omitted. The other way is to include 1964-I, reducing multicollinearity, in which case unlagged U is significant if most of the other variables are omitted, and so is σ^2—with the wrong sign! This should not, however, be taken very seriously. We shall consider briefly in Section IV below the significance of this model.

IV

The results reported in Table 7 above, at first sight frustrating because of the close correlation of the moments, in fact suggest something exciting: the possibility that the distribution of unemployment may be well described by a one- or two-parameter frequency function. If this were the case,

TABLE 6

U. S. STATES MODEL, QUARTERLY 1959-I TO 1967-I (EXCEPT EQUATION (1)):
REGRESSION EQUATIONS
(T-values beneath coefficients)

Const.	D_1	D_2	D_3	nU_0	nU_{-1}	$\dot{P}_0$	$\dot{P}_{-1}$	σ_0^2	σ_{-1}^2	$\bar{U}_0$	$\bar{U}_{-1}$	γ_0	γ_{-1}	D	R^2
5.39 (10.7)	−7.14 (15.1)	−5.63 (12.8)	−3.03 (8.26)	−0.935 (4.90)	0.846 (4.32)	−0.049 (0.18)	0.345 (1.97)							2.39	.912
3.73 (2.26)	−7.72 (8.37)	−5.78 (7.11)	−3.16 (4.95)	−1.05 (2.86)	1.23 (2.48)	0.402 (0.55)	0.940 (1.31)							1.93	.930
6.22 (5.58)	−9.09 (9.67)	−6.11 (8.86)	−3.49 (5.86)			0.482 (0.9)	0.323 (0.68)	0.226 (0.30)	1.68 (2.34)	−1.27 (2.86)	0.37 (0.72)	−0.47 (0.56)	−0.81 (1.39)	2.36	.970
5.98 (8.42)	−8.11 (13.1)	−5.44 (12.3)	−2.77 (7.08)			−0.089 (0.17)		−0.147 (0.22)	1.84 (4.0)	−0.843 (2.27)	−0.597 (1.13)			2.24	.955
6.45 (9.65)	−8.36 (13.4)	−5.64 (11.5)	−3.20 (7.51)						1.84 (4.47)	−0.974 (5.29)			−0.984 (2.43)	2.09	.967
5.50 (8.06)	−8.2 (9.6)	−5.58 (8.87)	−3.26 (5.92)				0.367 (0.75)			−1.05 (3.91)	0.931 (3.01)			2.02	.954

Notes:

$\dot{w}/w$ is measured by the percentage change in nonmilitary compensation of employees per man. Data required: compensation of employees in the private sector, compensation of nonmilitary government employees, and civil employment. Sources: $\dot{P}/P$ is the percentage change in the consumer's price index. Sources: *Business Statistics*, 1967, and *Survey of Current Business*, various issues.

$\bar{U}$, σ^2 and M_3 are the moments of state unemployment rates, 49 states. Up to May, 1963, observations refer to the central month of each quarter, thereafter to the first. Sources: *Unemployment Insurance Statistics*, various issues; *Employment and Earnings*, various issues. nU is the national unemployment rate. Sources: *Business Statistics*, 1967, and *Survey of Current Business*, various issues.

TABLE 7

U.S. MAJOR AREAS MODEL: MATRIX OF PARTIAL CORRELATION COEFFICIENTS
(From multiple regression with quarterly dummies)

	A. All Observations, 1959-I–1967-I						B. Omitting 1964-I				
	nU	$\overline{U}$	σ^2	M_3	γ		nU	$\overline{U}$	σ^2	M_3	γ
$\overline{U}$	.983					$\overline{U}$	.980				
σ^2	.447	.508				σ^2	.943	.963			
M_3	.011	.069	.884			M_3	.674	.754	.834		
γ	−.340	−.267	.612			γ	−.676	−.646	−.612		
$\Delta\overline{U}$			.422	.340	.263	$\Delta\overline{U}$			−.118	−.075	.011

Notes:
Variables as for Table 6, except for $\overline{U}$, σ^2, M_3 and γ, now calculated from quarterly unemployment rates in the 150 major labor areas.

1st Quarter: mean of January and March observations
2nd Quarter: May observation
3rd Quarter: mean of July and September observations
4th Quarter: November observation.

(to diminish seasonal effects of Christmas and summer vacations, etc.)
Source: *Area Trends in Employment and Unemployment*, various issues.

then we should be able to explore the transformation between excess demand and unemployment, and hence the frequency function for excess demand itself. Professor Brechling and I hope to investigate this possibility in the near future and we shall not discuss it here. There is, however, one point that may be made. Consider a value of $\overline{U}$, which will have been repeated many times in history, and the sectoral distribution associated with it. We should not expect the same sector to appear always with the same U_i value for given $\overline{U}$: industries disappear, new ones emerge, technology changes, demand changes. The automobile industry cannot have taken the same place in the distribution ever since 1861, the first year of Phillips' observations! Yet for a moderately stable aggregative relationship to have held, there must have been some structural stability. We may conjecture that the frequency function of excess demand may be stable, although the position of the individual sector cannot be. This conjecture appears more plausible when we further recall that the fitted macroequations for $\dot{w}/w$ totally neglect an insoluble nonlinear aggregation problem. Their relative success seems to imply some stability in the aggregation.

IMPROVING THE LABOR MARKET TRADE-OFF BETWEEN INFLATION AND UNEMPLOYMENT*

By CHARLES C. HOLT
Urban Institute, Washington, D.C.

I. *Introduction*[1]

The work on the Phillips curve has been predominantly empirical, but policy intervention designed to decrease both inflation and unemployment requires a better theoretical understanding of the determinants of this relation. This paper attempts to sketch out the basic labor market relationships that appear to the author to account largely for the Phillips relation[2] and then to consider the kinds of policy measures that this analysis suggests are relevant for moving the Phillips curve. The basic issue involved in the stability or instability of the Phillips curve when the inflation rate is constant is explored with a simple model in the appendix.

The emphasis in this paper is on the atomistic operation of the labor market because it appears sufficient to generate the Phillips relation, but we do not deny that union bargaining and price dynamics may also be involved to some degree.

This paper concentrates on conceptual issues. Continuing research will be devoted to the statistical measurement of relationships.

II. *The Structure of the Labor Market*

Since unemployment is viewed as a transitory state through which most workers pass rather than an experience suffered only by some particular kinds of people, it is relevant to examine the tremendous turnover of workers that keeps the labor market in a continual state of flux. Figure 1 shows various flows and stocks in the labor market that we shall discuss, giving rough numerical magnitudes.

The balance of this section is devoted essentially to an explication of this figure.

Starting at the bottom of the diagram, aggregate demand, which is influenced by monetary and fiscal policy, etc., when combined with product prices, determines the real demand for goods and services and, in turn, production which generates manpower requirements. If these exceed the manpower available from the current stock of employed workers, new job vacancies are created which add to the stock of unfilled job vacancies. If manpower requirements fall short of available manpower, layoffs occur and workers flow from the category of employed to unemployed, reducing one stock and increasing the other. Vacancies and unemployment influence hires and recalls.

Worker decisions to leave employment also account for labor turnover flows by (1) searching for better jobs while still employed and then quitting to change jobs without being unemployed, (2) quitting and engaging in unemployed job search,

---

* The research reported here was conducted under Contract 82-09-68-44 for the Manpower Administration, U.S. Department of Labor, under the authority of the Manpower Development and Training Act. Researchers undertaking such projects under the government sponsorship are encouraged to express their own judgment. The author expresses appreciation for helpful comments received from Harvey Garn, Ralph Smith, Richard Wertheimer and C. Duncan McRae.

[1] Due to space restrictions the original, "How Can the Phillips Curve Be Moved to Reduce Both Inflation and Unemployment?" could not be included in the *Proceedings*. Consequently this paper has summarized the conclusions of its technical analysis and has incorporated additional material, especially in Sections II, VI, and the appendix. The original paper will appear in *Microeconomic Foundations of Employment and Inflation Theory*, Edmund S. Phelps, ed. (W. W. Norton, forthcoming, 1969).

[2] For some other theoretical approaches see [8].

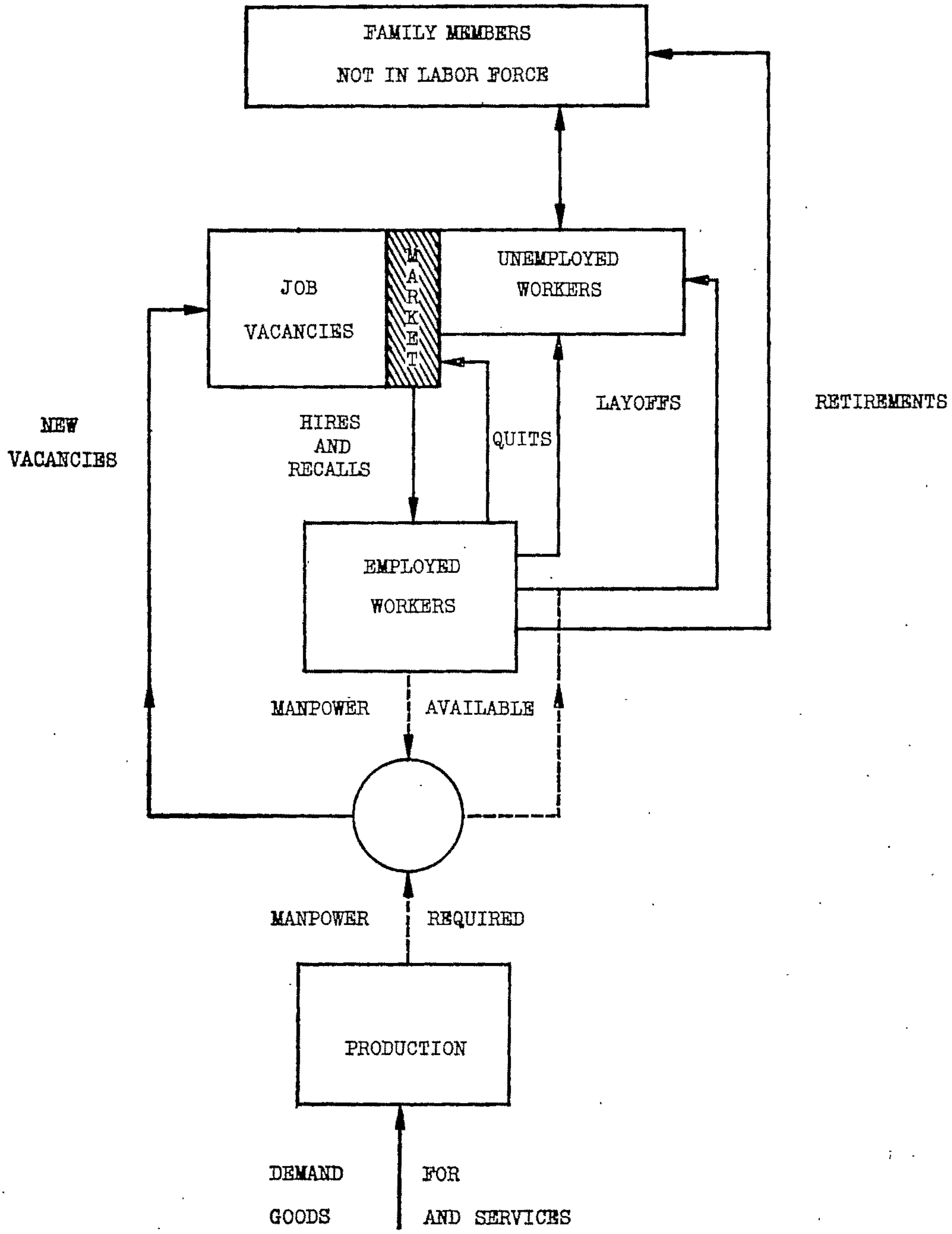

FAMILY MEMBERS
NOT IN LABOR FORCE
JOB
VACANCIES
MARKET
UNEMPLOYED
WORKERS
NEW
VACANCIES
HIRES
AND
RECALLS
QUITS
LAYOFFS
RETIREMENTS
EMPLOYED
WORKERS
MANPOWER
AVAILABLE
MANPOWER
REQUIRED
PRODUCTION
DEMAND
FOR
GOODS
AND SERVICES

FIGURE 1
FLOWS AND STOCKS OF WORKERS AND JOBS

and (3) retiring from the labor force. For various domestic and educational reasons family members are continually entering and leaving the labor force.

The following rough and rounded figures will give an impression of the magnitudes of the stocks, flows, durations, and probabilities involved in this system. In the United States currently there are about 80 million employed workers, 3 million unemployed workers, and probably a comparable number of vacancies, but data on this are inadequate. Employment durations for all accessions average roughly 2.7 years. Unemployed workers and vacancies wait roughly a month on the average before finding work or employees. The total flow of quits and layoffs amounts to roughly 30 million per year. Some workers even travel the hire and quit or layoff loop several times in a year. Turnover rates of the order of 3 or 4 percent a month account for this tremendous flow. The stock of unemployed workers is replaced every month on the average. Offsetting this flow from the stock of employed workers is a roughly equal flow of accessions. Employers have to recruit continually in order to hold a constant work force. The probability that an unemployed worker will find a job is roughly 20 percent per week and gives a long-tailed exponential distribution of unemployment duration.

Now consider a cyclical fluctuation in which production increases by an arbitrary amount, say 3 percent. Labor productivity of the existing work force then increases by about 1 percent through reducing slack time, etc. To recruit the larger work force that is now needed, new vacancies are generated, which when matched with workers, reduce the unemployment rate roughly 1 percent. The increased vacancies and reduced unemployment encourage family members to increase their labor market participation so the labor force increases by about 1 percent. Ultimately the vacancies are matched with workers so that the 3 percent increase in production leads to an employment increase of approximately 2 percent and an unemployment reduction of about 1 percent.

The average duration of vacancies is increased and the duration of unemployment is decreased. Workers encouraged by the increased vacancies and decreased duration of unemployment increasingly engage in on-the-job search for better jobs and quit when they find them or, alternatively, quit first and then search. Employers, finding recruiting more difficult and quit rates increased, respond by reducing the flow of layoffs. The total turnover flow changes little because the changes in layoffs and quits largely cancel.

Through the process of quitting and finding better jobs, workers are able to improve their money wages. Relatively low-wage employers suffer relatively high quit rates and long vacancy durations; and, in order to hold their work forces, are under increased pressure to raise money wages for their employees or to lower their quality standards in recruiting—this is equivalent to offering higher money wages. This process tends to force the laggard employers to change their wages in line with wage changes generated in the labor market.

Taken together these mechanisms can produce an upward creep of money wages.

The opposite effects, of course, occur for reduced demand. When unemployment is high enough, money wages fall steadily.

Thus, high turnover initiated by both employers and workers coupled with the heterogeneity of jobs and workers leads to the time-consuming search for new matches that accounts for stocks of unemployed workers and unfilled vacancies and continuing changes in the level of money wages.

III. *Relations between Unemployment, Vacancies, and Wage Changes*

The relationship that determines the flow of hires is of critical importance, because it, together with the turnover rate, determines the stock of unemployment. Let T_s be the mean search time that it takes for a particular unemployed worker to encounter a particular vacancy through a random search process and find sufficient mutual interest to schedule an interview. The average probability that such interviews will result in job offers and acceptances is P_{oa}. The probability per period of time that an unemployed worker will be placed is proportional to the number of vacancies that are available; thus the probability is $P_{oa}\mathcal{V}/T_v$. The total flow of new hires per period of time $\mathcal{F}$ depends on the above probability multiplied by the number of unemployed workers $\mathcal{U}$:

$$\mathcal{F} = \frac{P_{oa}\mathcal{V}\mathcal{U}}{T_v} \cdot \tag{1}$$

This says that the more efficient the search is in reducing the time between worker-vacancy interview encounters, the larger the number of vacancies and unemployed workers, and the higher the probability of successful interviews, the greater will be the flow of placements.

Solving for the stocks of vacancies and unemployed workers, we can see the determinants of their product:

$$\mathcal{U}\mathcal{V} = \mathcal{F}T_v/P_{oa} \, . \tag{2}$$

We see from Figure 1 that the stock of unemployed workers will continue to grow until the new hire rate is equal to the flow of quits, layoffs, and retirements; consequently the turnover rate determines $\mathcal{F}$.

Since the right-hand side of (2) fluctuates little over the cycle, we would expect to observe $\mathcal{U}$ and $\mathcal{V}$ moving cyclically in opposite directions; this has been observed; see (1) and (2).

The average durations of unemployment T_u and vacancies T_v are given by dividing the stocks by the flows:

$$T_u = \frac{\mathcal{U}}{\mathcal{F}} \quad \text{and} \quad T_v = \frac{\mathcal{V}}{\mathcal{F}} \cdot \tag{3}$$

Since $\mathcal{F}$ fluctuates little cyclically, we see that the duration of unemployment is high when unemployment is high, and similarly for vacancies. From (2) and (3) together we see that unemployment duration will tend to rise cyclically when vacancy duration falls and vice versa.

Since there are theoretical and empirical reasons [5] for expecting both unemployed workers and employers to have adaptive wage aspirations, we would anticipate that on the average when unemployment duration is long, workers would be willing to accept lower wages than when unemployment duration is short. Similarly employers will be willing to offer higher wages when vacancy duration is long. Consequently, when unemployment is low, workers tend to get increases in money wages when they go through the labor market but decreases when unemployment is high. Here we see that the turnover flow coupled with incremental wage changes which depend on the level of unemployment tend to produce a Phillips relation.

Elaboration of this model to include the responses of employers in making on-the-job wages changes is presented in [4].

If the labor market is segmented regionally, occupationally or by racial discrimination, relation (1) needs to be modified because some of the vacancies cannot reach some of the workers in their market search with the result that vacancies and unemployment are increased. If these segments are unbalanced so that vacancies are concentrated in some areas or occupations and unemployed workers in others, then unemployment and vacancies are increased still further.

An analysis of the effect of rigid compartmentalization is given in [5]. It shows that if, for example, the labor market were cut into nine equal and balanced isolated regions, total unemployment and vacancies both would be increased by a factor of three. If, in addition, there were an imbalance between regions in that the labor supply was tighter in some than in others, vacancies and unemployment would be raised still further.

The analysis of permeable barriers between regions, occupations, etc., has yet to be carried out and will be inherently more complex.

IV. *Existence of a Stable Phillips Relation under Steady Inflation*

Before we undertake an analysis of how the Phillips relation might be moved, we consider the argument presented by Friedman [3], Phelps [7], and others that no stable relationship exists between different rates of steady inflation and the level of unemployment. They concede that a dyamic relation exists and, undoubtedly would agree with Stein [9], that it is an important concern of dynamic stabilization policy.

However, their argument leads to the conclusion, which also is very important for policy purposes, that in equilibrium the level of unemployment cannot be affected by the rate of inflation and hence that we cannot reduce unemployment even by paying the cost of steady inflation.

We will present their argument and then give a counter example in the appendix to demonstrate that economies can exist which do not conform to their conclusion which purports to be generally true.

The Neoclassical Phillips Curve. The basic argument in its simplest form is that any steady rate of inflation would, after a time, be accurately foreseen by everyone and that adjustments would be made so that the inflation would have no effect on any real variables including production and employment.

Consider the situation in which prices are constant. Each individual faces an opportunity set that constrains his income, leisure, and time spent searching for better income opportunities. He maximizes his utility under these constraints and everyone else does likewise with the outcome that an optimal equilibrium level of unemployment U_e is reached.

Since any rate of steady inflation could be foreseen exactly, the level of unemployment would not be affected. Thus the only stable Phillips curve for the equilibrium case would be a vertical line at the unemployment level, U_e. See Figure 2.

However, it is recognized that a sudden increase in aggregate demand above the level required to maintain the steady rate of inflation could temporarily increase production, employment, and prices and reduce unemployment. Suppose, for example, that aggregate demand which initially grew at a rate consistent with steady increases in money wages at rate A suddenly started to grow so that wages rose at the higher rate B. The economy would move along the Phillips curve from point 1 to point 2. As the economy became acclimated to the new inflation rate, the Phillips curve would gradually rise and the unemployment rate rise until point 3 is reached where the "optimal" unemployment rate U_e is restored. Thus the Phillips relation is seen as relating unforeseen price changes and transitory fluctuations in unemployment. There are an indefinitely large number of Phillips curves, not a unique and stable one. Or put another way, the Phillips curve ultimately will adjust upward or downward in response to any steady inflation rate.

Since it may be permissible to neglect friction for long-run adjustments, this argument is quite persuasive.

A Search Model of the Phillips Relation. In order to challenge the universal validity of the above argument, we need a plausible counter example. This would then indicate that the applicability of the above argument to a particular case would depend on empirical knowledge of the "frictions" in the economy in question, and that its general applicability could not be simply assumed as has often been the case.

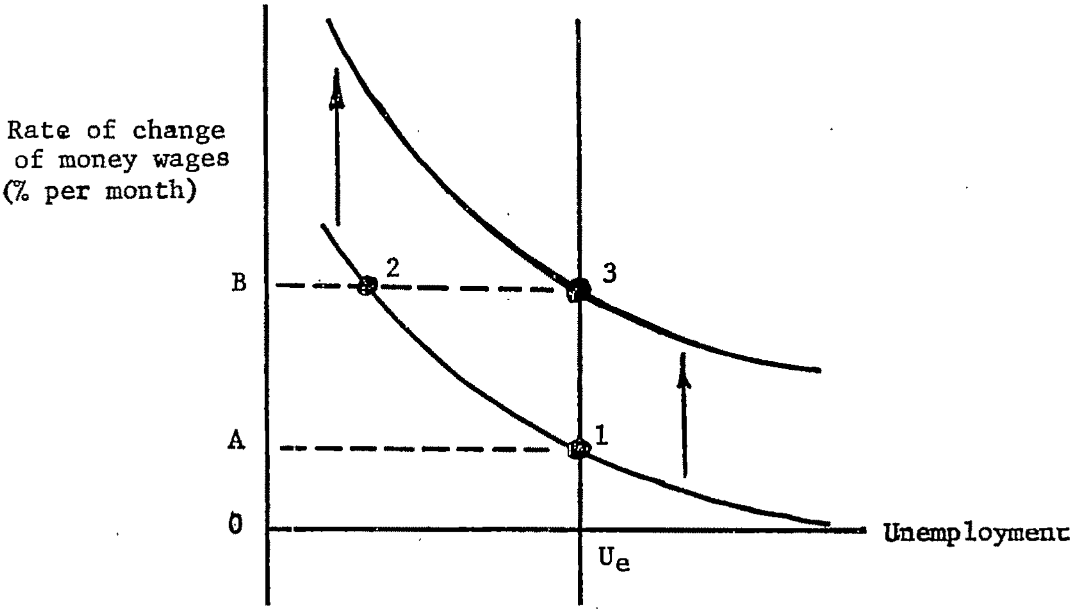


FIGURE 2
VERTICALLY SHIFTING PHILLIPS CURVE

The model that is presented in the appendix is intended to be simple rather than necessarily realistic. However, the behavioral relationships in the model could occur in an actual economy without stretching the imagination unduly.

This search model postulates that unemployed workers search the job market for the best paying jobs, but when workers are unsuccesful in attaining their wage aspirations, they gradually lower them as the search continues. The money wage aspirations are assumed to be only partially compensated for past inflation. Since long average duration of search accompanies high unemployment rates, wage aspirations fall considerably before workers are hired. Conversely, short search duration and low unemployment rates lead to relatively high hiring wages. This wage search phenomena coupled with a steady turnover of the labor force and no on-the-job wage changes[3] produces a unique stable Phillips relation that even under steady inflation does not drift up or down.

The key point is the determination of whether or not workers on the average make an exact compensation for past steady inflation. If they do, there is no unique steady-state Phillips relation, only a dynamic curve anchored to the equilibrium level of unemployment discussed above. If, however, there are some frictions in the system even in the steady state, then there will be a steady-state Phillips relation, not just a dynamic one.

In the model analyzed in the appendix, a parameter is introduced that indicates the degree to which workers compensate for inflation in their wage aspirations. At moderate rates of inflation and for the short run there are strong reasons to anticipate a less than complete compensation for steady inflation (or deflation). Even in the long run we are not safe in simply assuming that people living in a complex world of great uncertainty at the micro level will perfectly perceive and compensate for inflation in the aggregate wage level. The question is an empirical one and we badly need to develop and test a dynamic theory of market frictions and adjustment processes. Workers simply may not be able accurately to distinguish between wage inflation and changes in personal and general productivity.

---

[3] A more fully developed theory will, of course, provide for wage changes on the job, for wage changes from changing jobs without intervening unemployment, and for different aspiration levels for workers who quit and for those who are laid off.

Policy Implications. It is important that the manpower programs to be discussed in the next sections are equally suitable whether or not a steady-state Phillips curve is thought to exist. In one case, equilibrium unemployment is reduced and in the other, the Phillips curve is moved to the left.

The significant differences come in the implications for macro demand policy. If there is no steady-state Phillips curve, then we can have any one that we want including the ideal one that gives no price inflation at equilibrium unemployment. In this case the only question is how fast we want to achieve this complete solution to the inflation problem. High unemployment could give it to us quickly.[4] The optimal strategy would be to trade off the permanent inflation benefits against the temporary excess unemployment costs. Of course, if we really believed that everyone could perfectly compensate for steady inflation, we should have little policy concern at all about steady inflation, only unexpected inflation.

However, if there are indeed frictions in the economy that give a steady-state Phillips curve or even one that moved very slowly, a temporary increase in unemployment would give only a temporary reduction in inflation, not a permanet one, and we would be ill-advised to counsel increased unemployment to "break the inflation psychology"—unless we simply wanted to operate at a different static point on the Phillips curve.

The likelihood that there is a stable Phillips curve or at least one that responds very slowly to excess unemployment persuades the author that economists—and indeed national policy-makers should devote serious attention to the issue of how the Phillips curve might be shifted to the left.

V. *Policies for Moving the Phillips Curve*

From Section III we see that both vacancies and unemployment need to be reduced together if upward pressure on wages is to be avoided. Increases in aggregate demand offer an easy way to reduce unemployment, but the accompanying increase in vacancies might create an unacceptable inflation rate.

From equation (2) we see that decreasing the turnover flow $\mathcal{F}$ and the mean search time T_s would yield the desired reduction in both vacancies and unemployment. However, there are indirect effects that should be noted. Increasing search efficiency reduces T_s but through (3) we see that the reductions in unemployment and vacancies reduce their durations. This tends to raise the worker's acceptance wage and lower the employer's offering wage with the result that an interview is less likely to result in a placement; i.e., P_{oa} falls. In (2) this partially offsets the effect of improving search efficiency. Reducing turnover flow $\mathcal{F}$ would reduce the vacancies and unemployment but surprisingly through (3) would increase their durations. This would partially offset the impact on durations and the placement probability P_{oa} of the change in search efficiency that was just noted. It is clear that there are powerful feedback and interaction effects involved in the labor market system.

---

[4] Before buying this policy, we might want to examine the experience of the 1930's and determine whether the depression in fact gave us an improved Phillips curve.

Without going into detail, we can simply list the changes that could be influenced by policy actions to reduce unemployment (and vacancies) without increasing inflation:

1. Improving the information, search processes, and counseling of workers and employers would speed up the search[5] and reduce the average time T_s to arrange fruitful interviews.
2. Decreasing the segmentation of the market that is not necessitated in terms of production requirements, regionally, occupationally, racially, sexually, etc., will increase economic opportunities for both workers and employers.
3. Imbalances between labor market segments can be partially offset in the placement of government orders, stimulants to plant location decisions, the subsidy of geographic and occupational mobility by workers, and the restructuring of jobs to better fit worker abilities.
4. Interviews are more likely to result in placements, i.e., P_{oa} be increased, if the initial wage aspirations of workers and employers are set realistically close to the prevailing opportunities in the market. This should be improved by better information and counseling. Also, the faster workers and employers adjust their wage aspirations, the shorter their market searches will tend to be.
5. Employers can reduce their contribution to labor turnover by lowering their layoff rates through better sales forecasting, production, and employment planning, including seasonal smoothing, and through improved internal channels for promotion.
6. Workers can reduce their contribution to labor turnover by reducing their quit rates and reduce their unemployment by searching while still employed.

The high turnover rates[6] account for high training and other transition costs and undoubtedly could be greatly reduced before one need be concerned about excess rigidity in the economy. Making jobs more satisfying and improving the means of learning about jobs short of taking them could reduce turnover, but this important area is not well understood and requires much more research.

7. Since wage and price expectations can contribute to the inflation process, counseling efforts to reduce inflation expectations that are unrealistically high could be useful.

Although the above actions are oriented toward lowering unemployment and vacancies, they can be converted into reducing inflation by adjusting aggregate demand to change the vacancy-unemployment ratio.

VI. *Conclusion*

The foregoing policy recommendations are hardly surprising, but hopefully further quantitative research will contribute to our understanding of the relative benefits from the various alternatives.

Such manpower programs need to be coordinated with aggregate income policies, since manpower programs with no increase in demand are likely simply to redistribute jobs rather than reduce unemployment.

---

[5] With aid from electronic computers the duration of search by unemployed workers should be reduced, but in so doing it is important not to stimulate the turnover rate. If the offices of the U.S. Employment Service were kept open in the evenings, perhaps more of the job search could be done while workers are still employed but again the turnover caution applies. Efforts to reduce search costs need to be accompanied by efforts to make better placements that will have longer employment durations. For references on computer matching and other topics considered in this paper, see [5].

[6] Since 5 and 6 depend directly on private decisions, they can be influenced by government only indirectly. However, government leadership on the highest level may be helpful in getting employers to recognize that their individual decisions collectively have important national implications. Also, through counseling and computer aides, workers might be helped to find jobs that would satisfy them longer.

Manpower programs of the above type operate in the United States, but on a small scale. What are the arguments for increasing government resources in this area? The answer lies in the presence of externalities, equity objectives, institutional limitations, and economies of scale.

Most of the costs of labor turnover are not borne by the party that incurred them. When workers quit, employers must recruit and train replacements, often at higher wages. More than half of the quitting workers go directly to new jobs without any intervening unemployment to inhibit their moves. When employers lay off workers, they lose income while searching for new jobs. Government action is needed to minimize these serious external costs that are suffered involuntarily by both employers and employees.

Workers, in order to reach their maximum production potentials, particularly in a period of rapid changes in the structures and locations of jobs, require costly investments in education, training, job search, retraining, and relocation. Unfortunately, many do not have the money to make these investments themselves. Their employers may not make investments in training because the expected durations of employment are short.[7] Financial institutions typically refuse unsecured loans for such manpower investments by families because the risk is excessive, as there is no tangible asset that can be sold in case of default. Consequently, many socially and privately desirable human investments are not made, especially in disadvantaged workers, unless governmental action is taken. The inability to borrow to finance human investments may partially inhibit such investments for all except high-income families. The additional equity justification can be used for low-income families.

In the job and worker information area much of the cost is in obtaining the information in the first place rather than in its manipulation and dissemination. This is especially true where electronic computers are involved. There appear to be strong economies of scale that support the development of a small number of large efforts and in the extreme case lead to a natural monopoly as the most efficient form of organization. A strong case can be made for upgrading the resources, quality, scope, and computerization of the United States Employment Service.

Manpower programs offer the prospect not only of decreasing frictions and increasing efficiency of those directly involved but also may make it possible to increase production by increasing aggregate demand, decreasing unemployment, increasing labor market participation, and increasing the productivity of the whole employed work force—all of these are possible without contributing to inflation. Such programs seem to have significant potential for large-scale payoffs that can justify substantial increases in resources devoted to them. The 3 percent increase in production that was used as an example in Section III, amounting to perhaps $20 billion, may not be an unreasonable target that might be attained through the coordinated use of manpower and aggregate demand policies. A lower level of unemployment should have the additional benefit of stimulating productivity gains by increasing technological innovation and capital investment.

These policy implications are sufficiently important that continuing research is needed to test fully the view that the Phillips relation largely reflects the microdynamics of the labor market and to determine the specific effects of various manpower programs.

---

[7] Some training costs are, of course, shifted to workers in the form of lower wages during training periods, but some workers may not be able to afford such jobs even though their long-term payoffs are high.

APPENDIX

A Simplified Model of the Phillips Relation

The following is a simplified model of the labor market that can yield a stable Phillips relation under a constant inflation rate. It incorporates frictions of a type that may be important in operating markets and that may not disappear even in the steady state.

Each employed worker is assumed to have sufficiently little bargaining power to wield against the corporate bureaucracy of his employer so that he is unable appreciably to influence his own wage rate without leaving his present job and finding another.[8] This happens occasionally both voluntarily and involuntarily, but we assume that the turnover rate ($F' = \mathfrak{F}/\mathfrak{L}$ where $\mathfrak{F}$ is the flow of quits, layoffs, and retirements and $\mathfrak{L}$ is the size of the labor force) is constant.[9] When unemployment is high, layoffs are high and quits are low, and conversely. Hence the assumption of a constant turnover rate is reasonable and allows workers to quit more frequently when the inflation rate is high.

Once unemployed, a worker is assumed to have an acceptance wage that is the minimum that he will accept. Since employers want to keep their wage costs low, the offers presumably are close to this minimum or are at least related to it. Assume further that the money wage w_t was fully acceptable to the worker at the time that he started his previous job at time t and which lasted T_e time periods.

When he becomes unemployed at time $(t+T_e)$ and starts his search for a new job, he initially sets a high aspiration, reflected by the factor A, $A>1$. As his unemployed search duration T_u becomes extended, he lowers his aspirations exponentially at the rate D. He adjusts his wage aspiration for his own growth in productivity at rate p and for his estimate of the inflation at rate i that occurred in wages generally. Both the adjustments apply to the time span (T_e+T_u).

Thus the ratio of his acceptance wage at time $(t+T_e+T_u)$ to the wage that was accepted at time t is

$$\frac{w(t+T_e+T_u)}{w(t)} = e^{(i+p)(T_e+T_u)}Ae^{-DT_u} \tag{1}$$

The inflation rate is the difference between the rate of money wages generally, g, and the general increase in labor productivity $\bar{p}$:

$$i = g - \bar{p}. \tag{2}$$

The individual worker lacks accurate information on i, p, g and $\bar{p}$, so we add hats to indicate lagged perceptions of them and we introduce the factor C to reflect the extent to which he compensates for productivity and inflation. Ideally C would be unity. Using C and substituting (2), the first exponent would be

$$C(\hat{g} + \hat{p} - \hat{\bar{p}})(T_e + T_u). \tag{3}$$

Let g_j represent the long-term average rate of change in money wages for our typical worker whom we now designate j, assuming that he goes through a sequence of jobs. Thus his wage ratio shown in (1) could be expressed as

$$e^{g_j(T_e+T_u)} = \frac{w(t+T_e+T_u)}{w(t)} \tag{4}$$

---

[8] The cases of systems programmers and college professors changing jobs to obtain salary increases are relevant examples. Constant price agreements are not uncommon where costly gaming negotiations may be involved. Employers may consider it better to lose discontented employees individually on account of general wage policies rather than incur the larger costs of giving everyone wage increases which might become necessary if the wage issue were opened by negotiating with individual cases. When a new person is brought into an organization, there often is considerable flexibility on the wage paid to him depending on the particular job to which he is assigned.

[9] Usually turnover rate is expressed in terms of employees rather than work force, but the difference is small, and this slightly unorthodox definition contributes to the simplicity of the analysis.

Substitute (3) and (4) in (1) and take logs to obtain

$$(5)\quad g_j = C(\hat{g}+\dot{p}-\hat{p}) + \frac{ln\ A}{T_e+T_u} - D\left(\frac{T_u}{T_e+T_u}\right)$$

Now average this wage growth rate of the *j*th worker with all other workers to obtain an aggregate relation for the economy.

If people's perceptions about productivities are accurate on the average, $\dot{p}_j$ will equal $\hat{p}$, so in the interests of simplicity we will drop the $(\dot{p}-\hat{p})$ term.

Since (T_e+T_u) is the average turnover period from job to unemployed back to new job, it is related to the turnover rate.

$$(6)\quad F' = \frac{1}{T_e+T_u} = \frac{1}{\mathcal{E}/\mathcal{F} + \mathcal{U}/\mathcal{F}} = \frac{F}{L}$$

$\mathcal{E}$ and $\mathcal{U}$ are stocks of employed and unemployed workers respectively.

Also the fraction of time unemployed when averaged across workers is simply the unemployment rate

$$(7)\qquad U = \frac{T_u}{T_e+T_u}\,.$$

Substituting (6) and (7) in aggregated (5) yields

$$(8)\qquad g = C\hat{g} + F'\ ln\ A - DU.$$

Assume that the expectation of the general rate of price change at time *t*, $\hat{g}_t$, is simply its previous value g_{t-1}

$$(9)\qquad \hat{g}_t = g_{t-1}.$$

Substituting (9) in (8) and solving for the general rate of change in money wages yields

$$(10)\quad g_t + \frac{C}{1-C}(g_t - g_{t-1}) = \frac{F'\ ln\ A}{1-C} - \frac{D}{1-C}U.$$

This completes the formal development and analysis of the model. We now turn to its interpretation.

We could reasonably assume that the typical worker would recognize the real costs of search and unemployment, and, for a time, would avoid quitting his job even though he recognized that in continuing it his wages were not fully keeping up with inflation. He might also be partially trapped by the fact that some of his work experience might not be transferable and his productivity might be lower elsewhere. Also there is an element of uncertainty in changing jobs, and he is likely to have some aversion to risk.

Given the fact then that he has voluntarily accepted some real losses by failing fully to keep up with the inflation on his old job, it appears reasonable that he would set his aspirations to try to catch up on some of the inflation that occurred during the time span (T_e+T_u) but he may not think it feasible to catch up fully[10] and he accepts partial compensation and setting C less than unity.

Because each worker is likely to have difficulty in judging his own increase in productivity, much less judge how much of the general change in wages is due to inflation and how much is attributable to general productivity increases, he may well have some tendency to confuse his own productivity-based wage gain with a wage adjustment to compensate for inflation. This again would lead to undercompensation in setting wage aspirations for his next job.

In such a complex environment, workers may never learn to compensate fully, even when the inflation rate is constant.

With $0<C<1$, equation (10) constitutes a well-behaved stable Phillips relation that will tend to have clockwise loops. Here we can clearly see that high turnover F' and high aspirations A raise the Phillips curve. A fast rate of decline in aspirations represented by a high D lowers the relation. The closer the compensation C comes to unity the steeper the Phillips curve.

---

[10] He recognizes that employers in judging his productivity are influenced by his past wage.

For the case of perfect compensation where C equals unity, a case that certainly could occur, equation (10) after multiplying by $(1-C)$ becomes

$$(11)\qquad g_t - g_{t-1} = -D\left(U - \frac{F' \ln A}{D}\right) = -D(U - U_e).$$

This indicates that any steady inflation rate g is consistent with unique level of unemployment U_e. This corresponds to the neoclassical case discussed in Section IV. In this case the level of unemployment that corresponds to no change in the inflation rate is

$$(12)\qquad U_e = \frac{F' \ln A}{D}$$

which shows how turnover and aspiration parameters determine the level of frictional unemployment.

Equation (11) shows that unemployment above or below U_e leads to negative or positive accelerations of the level of money wages.

There may well be an interaction between the rate of inflation and the degree of compensation C, since the costs of not learning to make full compensation rise with the inflation rate. Thus a stable Phillips curve may persist only as long as inflation rates are moderately low. Very high inflation rates may boost C to unity and destroy the stability of the Phillips curve.

In any case, we would argue that we face an empirical question in deciding between the stable or the floating Phillips curve hypotheses and that it should not be assumed away simply by asserting that in the long run all frictions disappear. They may not!

This model is, of course, oversimplified, but hopefully it dramatizes the research issue that is involved.

REFERENCES

1. CHARLOTTE BOSCHAN, *Fluctuations in Job Vacancies—An Analysis of Available Measures* (N.B.E.R., forthcoming).
2. J. C. R. DOW AND L. A. DICKS-MIREAUX, "The Excess Demand for Labor: A Study of Conditions in Great Britain, 1946–56," *Oxford Econ. Papers*, Feb., 1958.
3. M. FRIEDMAN, "Comments," in G. P. Shultz and R. Z. Aliber, eds., *Guidelines, Information Controls, and the Market Place* (Univ. of Chicago Press, 1966).
4. CHARLES C. HOLT, "Job Search, Phillips' Wage Relation and Union Influence, Theory and Evidence," *Microeconomic Foundations of Employment and Inflation Theory*, Edmund S. Phelps, ed. (W. W. Norton, forthcoming, 1969).
5. ———, "How Can the Phillips Curve be Moved to Reduce Both Inflation and Unemployment?" *Microeconomic Foundations of Employment and Inflation Theory*, Edmund S. Phelps, ed. (W. W. Norton, forthcoming, 1969).
6. CHARLES C. HOLT AND MARTIN H. DAVID, "The Concept of Vacancies in a Dynamic Theory of the Labor Market," *Measurement and Interpretation of Job Vacancies* (N.B.E.R., 1966), pp. 73–141.
7. EDMUND S. PHELPS, "Money-Wage Dynamics and Labor-Market Equilibrium," *J.P.E.*, July–Aug., 1968, Part II.
8. ———, ed., *Microeconomic Foundations of Employment and Inflation Theory* (W. W. Norton, forthcoming, 1969).
9. HERBERT STEIN, "Economic Stabilization," *Agenda for the Nation*, Kermit Gordon, ed. (Brookings Institution, 1968).

THE NEW MICROECONOMICS IN INFLATION AND EMPLOYMENT THEORY*

By EDMUND S. PHELPS
University of Pennsylvania

It is notorious that the conventional neoclassical theory of the supply decisions of the household and of the firm are inconsistent with Keynesian employment models and with the post-Keynesian economics of inflation. Why should a fall of aggregate demand[1] reduce output and employment? Relentless application of textbook principles to either a competitive industry or a pure monopoly shows that output and employment in the corn industry will fall, given the technology and fixed factors, only if there is a rise of the product wage—the wage in terms of corn.[2] Why should a fall of demand raise product wage rates? The Keynesians, while postulating money wage "stickiness," never articulated a labor model in which the absence of automatic wage escalator clauses (tying money wage rates to the price level) could be explained. Further, it is widely agreed that product wage rates do not unerringly move "countercyclically" if they have any such tendency at all. What, then, is the appropriate theory of industry prices and outputs? Similarly, why should inflation (or unexpected inflation) be said in post-Keynesian economics to "buy" an increase of output and employment? Firms might be thought, neoclassically, to require lower product wage rates to produce more while households would be unlikely to respond with a corresponding increase of labor supply if real wage rates were to fall. It seems clear that macroeconomics needs a microeconomic foundation.

Now there has emerged the necessary kind of microeconomic theory of production, labor supply, wage and price decisions—a body of theory which is not restricted to conditions of neoclassical intertemporal equilibrium. The theory is different from some other efforts to buttress Keynes in that it sticks doggedly to the neoclassical postulates of lifetime expected utility maximization and net worth maximization, it makes no appeal to faulty perceptions (money illusion) and it does not plead that price setters economize on their decision-making time. Despite these abstractions, the new theory manages to deepen our understanding of traditional Keynesian economics and to adduce new hypotheses about macroeconomic behavior. It offers reasons why money wage rates are "sticky" in the face of price-level movements—why money wage rates (and thus prices) do not quickly respond by enough to keep employment and output at their equilibrium levels when aggregate demand changes.[3] It explains the stickiness of prices when money production costs move, so that real wage rates are not implied necessarily to rise when employment falls despite diminishing marginal productivity. The new theory generates a momentary steady-

---

* The author's work in this area is being supported by the Brookings Institution, for which he is preparing a monograph on public policy towards inflation and unemployment.

[1] By "aggregate demand" I mean (throughout) a schedule in the price-real income (or output) plane. This schedule may be vertical or, under Keynes or Pigou effects, negatively sloped.

[2] Aggregate demand can make absolutely no difference to any competitive firm's output except through the market-clearing prices and wage rates. In the monopoly case, there is the obvious elasticity qualification.

[3] Hence the previously advertised title of this paper, "Economics of the Absent Escalator."

state Phillips curve relation between the employment level and the rate of wage (or price) change; the rate of change of employment (or output) also frequently figures in the general Phillips relation. Finally, the crucial role which the new theory assigns to expectations, especially expectations of wage and price change, together with the notion of adaptive expectations has led most of the new theorists to the hypothesis that the Phillips curve is like a "predetermined" or "state" variable: today's Phillips curve may be largely inherited but tomorrow's curve will depend upon how the economy behaves today—in such a way, to be precise, that steady inflation will not "buy" a permanent (non-vanishing) reduction of the unemployment rate.[4]

The theoretical departure that is common to this otherwise neoclassical analysis is the removal of the Walrasian postulate of complete information. In the Walrasian economy, all transactions take place under complete information on the part of each buyer and seller about his alternatives. As a consequence, the Walrasian economy satisfies the two conditions for economic equilibrium.

The first of these conditions is that there is no nonprice rationing. There exists no buyer who, knowing he could make both himself and some seller better off by paying a higher price in return for more of the seller's product, is yet somehow frustrated from overbidding because the seller is (at least temporarily) ignorant of his bid. Similarly, no seller is frustrated form underbidding. This condition can also be expressed by saying that prices clear markets.[5]

The market-clearance condition means only that every price is market-clearing among the set of people who know the price. It means that every firm knows how much output those buyers who know his price are willing to buy from him at various prices he might set, and analogously for sellers to the firm. The second condition for equilibrium is that each household (and each firm if there are interfirm transactions) always knows the market-clearing prices of every firm now and in the future. There is perfect knowledge of current prices in the stringent sense that if a price were to change somewhere in the economy, every household would know the new price immediately. This requires there be no cost to learning or to advertising these prices.[6] The equally stringent condition that future market-clearing prices are known is best interpreted as a matter of perfect foresight or intuition.[7]

---

[4] This hypothesis, known variously as the "permanent unemployment thesis" or the "strong expectations hypothesis," should not shock scholarly economists. Fellner and Wallich argued it in this country more than ten years ago, and von Mises before them. The novelty in the work reviewed here in this connection is its explicitness about the expectational and learning mechanisms involved.

[5] That prices clear markets follows from the maximization postulate that no buyer or seller would knowingly pass up an opportunity for improvement together with the informational postulate that the firm knows its purchase and sale opportunities. In the case of a monopolist firm that sets price (nondiscriminatorily), market clearance means simply that the monopolist operates "on" his demand curve; he never rations buyers and he never produces in excess of quantity demanded at the price he sets because he knows his maximum sales at every price. Similarly, the monopsonist firm setting price operates "on" his supply curve. In the competitive case, where buyer and seller act as if the price necessary to effect a purchase or a sale were independent of the quantity purchased or sold, market clearing means that supply equals demand.

[6] It is not simply that the household thinks it knows all the prices and it happens to be right in the prevailing economic state; the weaker concept of non-Walrasian equilibrium involves the correctness of expectations.

[7] The analytical construct of a comprehensive futures market to determine all future prices in advance, so that future prices are placed on a par with current prices, falls to the ground in an economy with a positive birth rate. For the present members of the economy, to achieve economic efficiency, will want to be able to trade later with the yet unborn members of the economy who cannot as yet enter into futures contracts.

On the high ice-clad slopes of the Walrasian equilibrium economy, the question of a connection between aggregate demand and the employment level is a little treacherous.[8] Only one path of aggregate demand will produce the path of the price level and money-wage level that people foresee. But we may ask what difference it would make if people were spontaneously to foresee inflation (of prices and money wage rates equally) rather than a stationary price and wage level, so that aggregate demand must be steadily increased to validate these expectations. Would the inflationary scenario be one of higher employment than in the stationary-price story? The answer is that there is no necessary connection between the rate of anticipated inflation and the level of employment. Employment, investment, and other real supplies and demands will be invariant to the difference in the anticipated price trend if all foreseen relative prices, including foreseen real rates of interest, are likewise invariant. The two tools of fiscal and monetary policy give the government enough degrees of freedom to validate and sustain a "pure" or "neutral" inflation in which real interest rates are insulated from the expectations of inflation.[9]

I shall briefly discuss here the three types of disequilibrium models to be found in the new microeconomics under incomplete information. I start with the labor market side.

I. *Labor Markets and Money Wage Behavior*

Search Unemployment and Effective Labor Supply. A paper by George Stigler [13] has emphasized that labor markets are characterized by seriously incomplete information on the part of the worker concerning current wage rates elsewhere in the economy, so that a certain amount of "search unemployment" is normal. Armen Alchian has pointed out that, on a reasonable expectational hypothesis, the quantity of search unemployment and thus the level of employment will vary with aggregate demand through its effect on sampled money wage rates [1].[10] Specifically, an increase of aggregate demand will reduce search unemployment by causing the searcher to mistake a general rise of money wage rates for a lucky sampling of a high relative money wage offer which he believes he should accept.

Perhaps the simplest model would be a picture of the economy as a group of islands between which information flows are costly: to learn of the wage paid on an adjacent island, the worker must spend the day traveling to that island to sample its wage instead of spending the day at work. Imagine, only for simplicity, that total labor supply—the sum of employment and (search) unemployment—is a constant for every household, independent of real wage rates, expected real interest rates, and so on. Suppose, also, that labor is technically homogeneous in production functions and indifferent among the many heterogeneous jobs of producing a variety of products. Producers on each island are in pure competition in the labor market as well as in the interisland product markets. Each morning, on each island, workers "shape up" for an auction that determines the market-clearing money wage and employment level. To start with, imagine a very stationary set-up in which there is no taste change and no technical change, with constant population size.

---

[8] If one wishes to discuss money prices and inflation in the Walrasian economy, one needs to suppose that, even in the fully informed Walrasian economy, peoples' IOU's would not be completely trusted and that the government—or a few firms (banks) under government regulation—monopolize the manufacture of currency.

[9] Three qualifications to this conclusion may be mentioned: The expected capital loss on real money balances may substitute for taxation of income so that if taxes are not assumed to be lump-sum, there will be a substitution effect favoring employment and saving. Second, if legal or technological factors prevent money (or some components of money) from bearing interest, the opportunity costs of holding money will be greater under anticipated inflation and one could imagine that the resulting additional nuisance of managing transactions balances would shorten the workday a little or divert secondary household workers from the labor market. Third, there may be "distribution effects" on labor supply if the inflation was not anticipated as early as the signing of the oldest outstanding contract expressed in money terms.

[10] Mention should also be made of Leijonhufvud's new book [6] which views money wage stickiness in much the same way.

Initially, wage rates are moving as has been expected, and it is believed that unsampled wage rates (on other islands) are equal to the sampled (own-island) one. The economy is thus in a kind of non-Walrasian equilibrium in which wage rates are correctly guessed—though a change of some island's wage would not be immediately learned. For simplicity, suppose that money wages have been expected to be stationary. Now let aggregate demand fall. If the decline of derived demand for labor were understood to be general and uniform across islands, money wage rates (and with them prices) would fall so as to maintain employment and the real wage rate (provided that a new equilibrium exists). But suppose that workers on every island believe the fall of demand is at least partly island-specific due to their island's individual product mix. It is natural then to postulate that workers' expectations of money wage rates elsewhere (on other islands) will "adapt" less than proportionally to the unforeseen fall of sampled money wage rates. To the extent that the island-specific component of the wage change is believed to be enduring enough to make search for a better money wage rate seem worthwhile, the acceptance wage on each island will fall less than proportionally to product prices; some workers will refuse employment at the new (lower) market-clearing money wage rates, preferring to spend the time searching for a better relative money wage elsewhere.[11] Effective labor supply thus shifts leftward at every real wage rate, real wage rates rise and profit maximizing output and employment fall.

A pioneering paper by Charles Holt [4] and a closely related one by the present author [10] discuss the generation of a Phillips curve from models of imperfect information in labor markets. More about these models later. In fact, the faint shape of a Phillips curve relation between the steady unemployment rate and the rate of wage change can be seen to emerge from the ultra-simple island model. If the government were to manipulate aggregate demand to keep the average money level constant at its new lower level, the search unemployed would be disappointed at finding money wage rates equally low elsewhere and would hence revise downward their expectations of the mean wage elsewhere relative to sampled wage rates; search would become less attractive and effective labor supply would shift rightward.[12] To prevent the market-clearing employment rate from rising, therefore, the government would have to continue to reduce money wage rates by contracting aggregate demand (or by holding the aggregate demand function steady if aggregate demand is independent of the price level). This action would be effective on the hypothesis that every unexpected decrease of sampled money wage rates produces a less than proportional decrease of expected money wage rates elsewhere. Then each day, there is some rate of decline of money wage rates that would keep sampled money wage rates equal to a constant fraction of expected money wage rates elsewhere. Thus some continuing decline of money wage rates (of the right magnitude) accompanies the maintenance of the specified volume of search unemployment.[13]

---

[11] I assume that workers differ in age, and hence differ in their appraisal of the lifetime gain from a specified expectation of wage-rate improvement, or that workers differ in the "adaptability" of their wage expectations, so that each island's effective labor supply curve is upward sloping.

[12] Is there a nonsteady-state path of return to steady-state equilibrium along which the expectations of wage rates by searchers are continuously validated? The answer is apparently no if expectations are identical across workers (and not only in that case). To confirm the expectations of the hopeful searchers, money wage rates must rise. To confirm the expectations of nonsearchers with like expectations (but older people), money wage rates cannot rise by as much—for there was an expectation of a change in wage relatives as well as the possibility of some return to normal money wage levels. (There might exist odd configurations of heterogeneous expectations which could be validated.) The nonsteady-state equilibrium paths of my recent paper [10] could be viewed as paths which make the wage rates move in conformity with the expectations by the employed of the trend of wage rates elsewhere.

Clearly, the "required" rate of decrease of money wage rates is larger the greater is the shortfall of actual wage rates from expected wage rates. It is also true that the volume of search unemployment is larger the greater is this shortfall. Hence we deduce a Phillips-like relation between the steady level of unemployment and the algebraic rate of increase of money wage rates. In this relation, the expected long-run trend rate of money wage increase figures as a parameter. If workers look backward and see that money wage rates are steadily falling and adapt their expectations of the general wage trend accordingly, an ever accelerating rate of decrease of wage rates will occur if the search unemployment level is maintained.

In the above story, every steady state of positive unemployment is one of disequilibrium in the sense that sampled wage rates are continually and systematically different from (less than) what they were expected to be.[14] Steady-state equilibrium occurs only at zero unemployment. To escape this implication, it is necessary to introduce structural change, like "real" microeconomic product-demand shifts, relative cost shifts, or perhaps (simple) population growth.[15] Then the islands where money wage rates are above the average money wage rates expected elsewhere will be numerous enough relative to the islands where wage rates are below expected wage rates elsewhere that the equilibrium steady unemployment rate (at which, on average, money wage rates move as expected) will be positive. There will be enough job "vacancies"—as defined by the quantity of labor that would be demanded at expected mean wage rates elsewhere minus actual employment where wage rates are "high"—in relation to the quantity of search unemployment that a kind of equilibrium "in the large" is possible in which, while individual searchers and nonsearchers may be disappointed or delighted, the mean rate of change of money wage rates is equal to the average expected trend rate of change.[16,17] In the model so extended, "overemployment" is possible. It results when, starting from the equilibrium level of search unemployment, money wage rates rise faster than expected. Such unexpected rises induce some of the search unemployed to stop search earlier, to accept employment at the sampled wage rates (and some employed to postpone search)—if, as hypothesized before, every unexpected wage increase produces a less than proportional increase of mean expected money wage rates elsewhere.[18]

---

[13] A rigorous argument that the rate of wage decline is constant—or simply asymptotically constant—is undoubtedly difficult.

[14] If the expected long-run trend rate of money wage change is, say, 4 percent, a small enough steady unemployment rate will be associated with rising money wage rates; but they will be rising at less than 4 percent, so that the same overestimates of wage rates elsewhere will exist and disappointment will occur.

[15] G. C. Archibald has been looking into the kinds of continuing sequences of structural change required to keep the steady-state equilibrium unemployment rate positive.

[16] If every area is searched (in equilibrium), the equilibrium so defined seems also to be characterized by equality of mean wage levels elsewhere with the actual mean wage level, on the average over workers. If easterners do not search the West, believing wage rates drastically lower there, then that characterization need not hold (whether or not westerners search the East).

[17] Note that in the kind of economy I have been sketching, wage rates will be high and falling where vacancies are defined to be present, while in sectors where wage rates are below expected wage rates elsewhere (the current loci of the unemployment) wage rates will be low and rising. These rates of change, I believe, need not characterize more thoroughly non-Walrasian markets where each firm is an island.

In summary, the search unemployment model suggests a wage change equation in which the Phillips relation is one element. The rate of wage change is connected to the level of the unemployment rate, the rate of change of employment and the expected trend rate of increase of wage rates —the latter entering the equation with a unitary coefficient. The first of these latter variables reflects the expected rate of wage improvement from search, and the second of these variables measures the rate of change in this expected improvement.[19]

Holt's paper presents a richly detailed model of job search, including on-the-job search by employed workers. In his model, the rate of overall wage increase is a weighted average of the rate at which unemployed persons reduce their acceptance wage (as they learn about the wage distribution they are sampling from) and the rate at which employed persons improve their wage rates by moving from low-wage to high-wage jobs. The smaller the steady-state unemployment rate, the larger is this weighted-average rate of wage change. The way in which a rise of general wage rates reduces the unemployment rates is also discussed. The effect of expectations of general trend rates of change in wage rates seems to be abstracted from in Holt's paper.

The present author's recent paper on wage dynamics and labor market equilibrium emphasizes the economics of wage setting and recruitment by the firm which is in an incompletely informed labor market. Here each firm is a kind of island having transient, dynamical monopsony power at each point in time: a rise of its wage offers relative to those elsewhere will increase the speed with which its employment roll will increase. Each firm will pay a higher wage relatively to its expectations of money wage rates elsewhere the smaller is the expected unemployment rate and the larger is its own expected job vacancy rate.[20] If expectations of going money wage rates elsewhere are "static," then, when all (or most) firms pay more than they believe wage rates elsewhere to be—a situation of "generalized excess demand" in which vacancies are high and unemployment low—recruitment results are disappointing, there is learning of actual wage rates elsewhere, and consequently general wage rates float upwards. It is argued that unemployment and vacancies together determine the rate of change of employment as well as that of money wage rates, given expectations of the long-term trend rate of change of money wage rates elsewhere, so that the unemployment rate and the rate of change of employment are interpretable partly as joint proxies for generalized excess demand in Phillips curves. But even with zero generalized excess demand, it is maintained, money wage rates will change at a rate equal to the expected rate of change of average money wage rates elsewhere. On the hypothesis of adaptive expectations, faster growth of aggregate demand cannot achieve a permanent reduction of the unemployment rate below its equilibrium steady-state level. Further, faster productivity growth would not improve (and might worsen) the equilibrium unemployment rate.

---

[18] Just as the average length of search is shortened, one might expect that fewer people would quit in order to search when money wage rates are rising unexpectedly fast. It seems to be the case, however, that quit rates are higher the smaller is the unemployment rate. To accommodate that generally accepted hypothesis, it is apparently necessary to add that quit rates depend upon the unemployment rate because the latter reflects expectations of the time required to find employment. This introduces nonmarket-clearing factors.

[19] I should mention two weaknesses, neither necessarily serious, which are peculiar to the above search-unemployment model. First, a fall of aggregate demand may fail to produce the expectation of finding better relative money wage rates elsewhere, and thus fail to increase search unemployment, if workers observe that the cost of living has fallen in proportion to sampled money wage rates (or in greater proportion) and if they take those consumer prices to be some indication of general wage rates. (This suggests that price-level stickiness has a novel role to play in search unemployment.) Second, if an unforseen inflation is marked by an abnormally low expected real rate of interest, the search-shortening effect of the unexpectedly high wage rates sampled could, in principle, be offset by the reduced discounting of the future wage gain expected from continued search.

[20] Measured at the expected going wage.

A new paper by Dale Mortensen [8] combines and further develops some of the notions of Holt and myself, obtaining a logically complete and rigorous model of the labor market. This important paper sustains and illuminates most of the foregoing propositions about wage and employment dynamics in models of imperfectly-informed labor markets.

"Wait" Unemployment, Nonmarket Clearance and Stochastic Factors. A recent paper by Donald Gordon and Allan Hynes [3] explores the implications of probabilistic demands (of unknown mean frequency) for a service in a statistical-decision-theoretic model of optimal price setting by the seller. The paradigm there is the pricing (or rental-setting) problem of the manager of a group of apartments. But Gordon and Hynes rightly insist that such a model has much to offer as a description of labor markets as well.

The authors maintain that every wage contract between employer and employee is, at least implicitly, a contract for some finite duration of time. The labor supplier whose service is, for the moment, not in demand at his standard wage will, if he reduces his wage enough to become employed, take the risk of having to reject an offer of employment at his standard wage in the near future. The supplier who faces a stochastic demand will want to set his wage high enough that, if he has correctly estimated the (stationary) demand distribution, he will be intermittently unemployed due to stochastic fluctuations in demand. Such a model certainly sheds light on the sporadic unemployment of artisans, lawyers and actors for whom a contract now will preclude new employment for some time.

When there is an increase of the mean frequency of demand, due to a change of aggregate demand, say, idle suppliers will accept employment at standard rates. But as the unusual frequency of demands persists, suppliers will adaptively revise their expectations of the mean demand frequency and raise their standard wages accordingly. Employment in excess of the equilibrium level, at which mean demand frequencies are correctly estimated, thus generates a succession of wage increases (in excess of the normal trend rate of increase). But as suppliers find that steady escalation of their fees is not enough to return them to their objectively optimal average idleness, wage rates will accelerate. Gordon and Hynes conclude that their model too denies the possibility of a stable trade-off between employment and inflation.

Probably the idea of a buffer stock of free time can be applied more broadly to labor markets. The employer who cannot recruit and train labor costlessly may be willing to employ—i.e., pay wages to—a buffer stock of idle or near-idle workers to insure his being able to meet randomly high demands for his products. Further, the employer who faces a stochastic supply of labor—at each wage, how many employees will quit tomorrow and how many unemployed will seek and accept jobs?—may find it optimal, frequently or even normally, to set his wage *above* the market-clearing level.[21]

The Neoclassicial Speculative Labor Supply Model. The most neoclassical of the disequilibrium models drops the postulate of perfect foresight, while retaining the other informational postulates of the Walrasian system. All current prices are market clearing and are universally known.

Consider an economy initially in Walrasian equilibrium: thus far the future has been correctly guessed. For simplicity, suppose that the expected price trend has been level. Now let aggregate demand increase so as to produce an unexpected rise of market-clearing product prices. If we abstract from speculative inventory accumulation and the use of goods in process in some industries, firms' production and employment will be unchanged if current real wage rates are unchanged (provided there is no change of the demand elasticities perceived by firms). The unforeseen price rise can increase the willingness of firms to hire labor only by somehow reducing market-clearing real wage rates. Now why should this occur? Why should the behavior of labor supply cause money wage rates to rise less than proportionally to the rise of the current price level?

The explanation offered is in part that workers will speculate on the future return of the price level toward its original level. Workers will "adapt" their expectations of the future price level, on this hypothesis, only partially in response to the unexpected rise of the current price level. Hence, if workers expect to spend any increment of money wages partly on future consumption goods, a rise of money wage rates in proportion to the current price level would make additional labor supply more attractive. Thus the labor supply curve shifts rightward at every real wage rate. Therefore real wage rates fall and employment is increased.[22]

This rationale for money wage stickiness was briefly discussed in 1952 by James Tobin [15] though the entire focus there was on the possibility of "underemployment" arising from an unforeseen fall of the general price level. Robert Lucas and Leonard Rapping [7] have now produced an intertemporal lifetime utility-maximization model of household labor supply, involving present and future consumption and leisure which rigorizes this kind of analysis.

Out of the Lucas-Rapping model comes not only a theory of money wage stickiness but a Phillips curve relation between the rate of price change and the employment rate. If aggregate demand is manipulated by the government in such a way as to keep the price level constant at its new and higher (and unforeseen) level, workers will (on the above adaptive-expectations hypothesis) successively revise upwards their expectations of future consumption-goods prices; these revisions will gradually reduce labor supply (at every real wage), thus driving money wage rates up and driving profit maximizing output and employment back toward their steady-state equilibrium levels.[23] But there is some rate of continuing unforeseen price increases, a rate which is able to keep the expected future price level equal to a constant fraction of the actual price level—if each unforeseen increase of the current price level induces a constant and less than proportional increase of the expected price level—which can thus induce households to maintain their labor supply at the given "overemployment" level. But note that the required rate of inflation is constant rather than ever accelerating, only if what might be called the "expected normal long-run trend rate of price change"—set equal to zero in this exposition—fails to adapt to the actual trend of the price level. In the Lucas-Rapping formulation of the price-change (or wage-change) equation, the expected long-run trend rate of price enters with a unitary coefficient.[24]

---

[21] Then it is not just the worker's expectation of money wages elsewhere but also his expectations of the number of job vacancies and the number of unemployed looking for them that will determine his reservation wage and quit decision.

[22] Even if workers did not save, one could build a model arriving at the same result in which employment during the current period produces output during the current period while wages are received at the end of the period and spent over the next period. But it is hard to believe that the lag of wage payments behind wage accruals is long enough for this mechanism to be effective in stimulating employment; and if the average lag of output behind the application of labor input is equally long one would need to introduce asymmetrical expectations as between the individual firm's own future price and the future general level of prices.

II. *Price and Output Dynamics*

In the above models, there was postulated to be complete information on the part of all buyers about all goods prices. If a firm were to change its commodity price, the full effect of this change on the quantity demanded of the firm's commodity would be felt instantaneously. In that kind of model, inducing the firm to produce more—to move down its demand curve expressed in wage units as it were—requires a fall of product wage rates or, more loosely, of real wage rates. No one believes that production fluctuations can be accounted for completely, if at all, by such countercyclical real wage rate movements. More sophisticated views of pricing which dispense with various Walrasian informational postulates can avert the implication that real wage rates must so behave if aggregate demand changes are to have output effects.[25] They can also rationalize a kind of excess capacity phenomenon as a normal accompaniment, like unemployment, of the non-Walrasian economy.

Customer Search, Optimal Pricing, and Output Supply. Recent work by Sidney Winter and myself [11] analyzes a model in which, because of sluggishness in the diffusion of information, the firm finds itself at every moment having at least transient monopoly power: it will not instantaneously lose all its customers if it raises its price nor gain the whole market if it cuts its price.[26] Yet over time the effect of such a price variation will be more appreciable as information on price differentials becomes more diffused through the market. I shall restrict attention here to the special case in which the firm is asymptotically competitive; it knows that setting a price permanently above the expected "going price" in the economy for comparable goods would ultimately cost it its entire market share, while setting a price permanently below the going price would ultimately capture a huge number of customers. In such an economy, we can imagine a non-Walrasian equilibrium in which, on the product side, for every firm, the inflow of newborn customers just matches the outflow of customers; the firm believes therefore that its price is just competitive, being equal to prices charged elsewhere; likewise, its customers believe they are being charged the going price elsewhere.

---

[23] One might ask, is there some nonsteady-state path of employment back to steady-state equilibrium, a path along which new and current expectations are validated? No, not generally, I should think, for the path that ratifies price expectations need not presumably ratify expectations of money wage rates.

[24] It must be emphasized that the unforeseen rise of the price level causes workers to believe that an additional dollar of money wage earnings offers increased command over future goods only if the unforeseen price increase causes a rise of the expected real rate of interest—the money rate plus the expected rate of decrease of future consumption-good prices. The latter is assumed to increase when there is an unexpected increase of the current price level; but this assumes that the regressive effect—future prices are expected to regress to their expected long-term trend—overcomes the extrapolative effect—the long-term trend rate of increase may be adapted upward. Is that assumption generally appropriate? As for the money rate of interest, there is no reason to suppose that it invariably falls by the amount of the expected rate of decrease of future prices. But what of an inflation that is triggered by open market purchases or credit expansion? A considerable literature suggests that inflationary episodes are frequently associated with a depression of expected real rates of interest. Perhaps Lucas-Rapping are right for the opposite reasons: unexpected price increases may be typically extrapolated, money interest rates may typically rise (if at all) by less than the expected rate of future price increase and the income effect of the resulting fall of expected real interest rates induces an increase of labor supply.

[25] In a manner of speaking, the derived demand for labor can be shown to shift like the effective supply of labor.

[26] The analogy to the Phelps-Mortensen attention to the dynamical monopsony power of the firm in labor markets should be clear.

There are two obvious theorems about "markups" (over marginal cost) in that equilibrium. The first propostion is that the firm does not fully exploit its transient monopoly position. It produces beyond the point where marginal cost equals instantaneous marginal revenue because it estimates that a temporary rise of price, while bringing a larger cash flow for the immediate future, would be offset in the future by the resulting erosion of its market share. The second proposition is that the firm produces less than that output rate at which its marginal costs would equal the equilibrium going price. To obtain the larger cash flow which greater sales at the going price would seem to offer, the firm would have temporarily to reduce its price in order to attract additional customers. Provided the expected real rate of interest is positive, there is some gap between marginal cost and price which is small enough that the present sacrifice of cash flow entailed by a small temporary price cut (remember the firm is already producing too much from a myopic view) is not worth the discounted future increase of cash flow which the resulting customer gain would bring.[27]

There is a kind of excess capacity in this equilibrium state. If we think of the array of "machines," each of which can produce a unit of output but which generally differ in their labor requirements, then some of the more labor intensive of these machines will be idle even though, at the going product wage, some revenue would be left over as rent for these machines if their output could find a buyer without the aforementioned costs. There may be excess capacity in another sense: it can happen that output increases are obtainable from an increase of aggregate demand without any decline of product wage rates.[28]

To see that in principle real wage rates can move procyclically, we note that, given the parameters controlling the expected rate of response of customer flow to a change in the firm's price, the firm's optimal price is homogeneous of degree one in the height of its instantaneous demand curve, the money wage rate and the expected going price elsewhere.[29] Its optimal output is homogeneous of degree zero in these variables. Now suppose that an increase of aggregate demand produces an isoelastic upward shift of the firm's demand curve and that money wage rates happen to rise in the same proportion. If the firm's expectation of the going price at competing firms is unchanged or is revised upwards in smaller proportion to the rise

[27] The same propositions can be deduced if the firm is in an analogous position in the labor market, having at every moment some transient monopsonistic power yet being an asymptotically competitive buyer of labor.

[28] Nevertheless, there need be no excess capacity in the sense of depressed output at the equilibrium employment level. Decrees to cut price relative to marginal costs would increase national output only if an increase of real wage rates would increase the quantity of labor supplied.

[29] For this exercise, I postulate a Walrasian and competitive labor market in which the firm is a wage taker.

of its own demand, then it is possible that the firm will want to raise its price less than proportionally to its instantaneous "demand price," meeting the additional quantity demanded with greater output (and greater employment). To raise its price in full proportion to the rise of the demand price at the intitial equilibrium output would be to neglect its belief that such a price increase would worsen its competitiveness.[30] The disequilibrium result, then, is a rise of real wage rates and an increase of output and employment.[31] In the symmetrical situation of a fall of demand and of money wage rates, producers let their markups increase, failing to realize that they must cut price further if they are to maintain their competitiveness. Thus the price level may fall by less than the money wage rate level despite the fall of output and employment.[32]

A Phillipsian relation between the output level (relative to capacity) and the rate of price increase results if each firm continuously adjusts upwards its price as it learns that it is not experiencing a net loss of customers from its higher price and as money wage rates keep pace with the general price level. The now familiar qualification must again be made: the long-run trend rates of increase of money wage rates, of demand, and of the going price elsewhere expected by the firm figure as parameters in any such steady-state Phillips relation. As these expected trend rates adapt to perceived trends, the Phillips relation floats upward.

Inventories, Excess Capacity, and Stochastic Demand. Still another model of economic fluctuations in output emerges from recent analyses of optimal pricing and/or production by the supplier of a good under conditions of uncertainty about future price or demand. The uncertainty attaches to the profitability of leaving some output unsold in the present when a choice must be made between selling the output in the present or in the future.

Edward Zabel [17] has studied a model of production and sales by a perfectly competitive firm that can store its output under increasing marginal storage costs. In the simplest version, the firm has a stationary subjective probability distribution of the market price in each future period. The occurrence of an abnormally low price will normally cause the firm to sell relatively little and to increase its inventory. If the money wage facing the firm is constant and if the firm's inventory holdings were not already large, the increase in their quantity will reduce the firm's optimal output; for with rising marginal holding cost, it is more likely now than last period that any increase of output (relative to last period's) would end up being sold next period for the mean expected price—rather than being held for a somewhat higher expectation of revenue on chance of an abnormally high price subsequently—since marginal holding costs of inventory have increased with the increase in inventory held. Whether total output in this competitive economy will fall depends upon the response of money wage rates. If we postulate a positively sloped labor supply curve against money wage rates (which is plausible since expected mean future prices levels are constant, though the expected real interest rate has increased since Zabel holds the money interest rate constant), aggregate output and employment will fall. In a recent note, I looked into the question of whether money wage rates could be depressed proportionally more than the current price of goods, so that "measured" real wage rates are not generally implied to move countercyclically [9]. This appears to be a possibility. Note finally that adaptations by firms of their expectations of the mean price when actual prices continue to deviate from the expected mean price will produce a Phillips phenomenon —with the usual qualifications about the "stability" of the trade-off.

---

[30] Formally, the output effect in this situation can be viewed as the result of a *ceteris paribus* fall of the prices which the firm believes other firms to be changing. Such a fall will cause it to reduce its markup, being grateful for whatever additional output is demanded as a result.

[31] D. H. Robertson seems to have had the clue to this process when he wrote: ". . . the stimulus of rising prices is partly founded in illusion . . . [the business leader] is spurred on . . . by imaginary gains at the expense of his fellow business men. It is so hard to believe at first that other people will really have the effrontery or the good fortune to raise their charges as much as he has raised his own" [11].

[32] This result is a possibility, not a necessity, in the model. Further, I believe that it is a less likely result if the firm has transient monopsony power. In an unforeseen inflation, the tendency to underestimate going wage rates may tend to dampen the firm's wage increase relative to its price increase.

Previous work by Oscar Lange [5] and by Robert Clower [2] was addressed to the problem of the ignorant monopolist facing a stationary demand curve. The aforementioned paper by Gordon and Hynes analyzes the monopolist who faces a stochastic demand curve. Their paper fits the problem of the owner of an apartment building who must offer one-year leases on his apartments. Simplifying somewhat, they imagine that the owner sets a rental on vacant apartments each day or period. The owner does not aim to set the rental so low that there would be a high probability of no vacancies each day. A higher rental is superior for the usual buffer-stock reason.

An increase of mean demand will at first reduce the vacancy rate. Rentals on vacant apartments are sticky because the increase of mean demand is mistaken for a random deviation. As this experience with the below-equilibrium vacancy rate continues, the owner will begin to "adapt" his subjective estimate of the mean demand at his current rental, raising his rentals accordingly. If aggregate demand is steadily increased so as to permit the below-equilibrium vacancy rate to continue, the upward adjustments of apartment rentals continues as owners grope for the mean of the distribution. This is the temporary Phillips relation at work. But owners eventually learn that they must raise rentals faster than they have been doing if they are to succeed in restoring their vacancy rates. Gordon and Hynes declare that there cannot be a stable law of disequilibrium price dynamics for the same reason that one cannot have a stable law for use in predicting individual stock price movements. When others learn the law; it is no longer descriptive. Their point goes deep but just how deep we cannot pursue here.

True "User Cost" and Neoclassical Speculative Capital Utilization. Following Keynes and Sidney Weintraub [16], there has been a revival of interest in the consequences of (true) "user cost." That term refers to the fact that some machinery deteriorates faster the greater the speed with which it is run and the greater the number of shifts per day (because maintenance opportunities are thereby reduced). Paul Taubman and Maurice Wilkinson [14] have looked into the consequences of user cost for the optimal output and investment decisions of the competitive firm, generally under static expectations.

The concept of user cost together with adaptive expectations may be capable of offering an additional reason why unforeseen price movements can cause output changes despite unchanged product wage rates. Let all product prices and money wage rates rise unexpectedly and in the same proportion. Suppose that prices and wages are adaptively expected to regress gradually in the direction of their previous level. Consider now the producer who can neither sell his existing capital nor buy additional capital. It is then obvious that the firm will want to add shifts and speed up its machinery since marginal money quasi-rent from so doing increases in proportion to current prices and wage rates, while the money shadow-price of capital utilization, which reflects future expected quasi-rents, increases in smaller proportion.[33] Thus the process of unforeseen inflation can produce a rightward shift of output supply and labor demand through user cost.[34] But, once again, increasing inflation is presumably necessary to maintain the disequilibrium.

III. *Conclusions*

There is a common thread running through all these models. Each one postulates that the actors of the model have to maximize as best they can under incomplete information about the future or even about the present situation outside their respective sampled domains. Lacking the desired knowledge, the actors (at least implicitly) form expectations of the state of the economy—over space and over time. The supply prices of outputs and of labor services and the demand prices for labor are linear-homogeneous in known and expected prices (including expected mean demand prices in the stochastic case)—present and future; and quantity decisions are homogeneous of degree zero in these variables. On adaptative or more general error-correcting expectational hypotheses, a change of aggregate demand alters the relations between known or sampled prices and expected prices. The implied alteration of expected relative prices—of expected wage rates elsewhere relative to sampled rates, of expected mean future demand prices relative to current demand prices, of expected real rates of interest, etc.—causes a change in quantity decisions, hence employment and output. If public policy maintains a disequilibrium in which expectations on average are systematically in error, learning causes expectations to be revised—and with them the supply and demand prices for labor services and the supply prices of outputs. It seems to be widely believed that maintenance of the disequilibrium, and thus the regeneration of these expectational errors, requires or would cause explosiveness in the rates of change of actual prices.

References

1. A. Alchian, "Information Costs, Pricing, and Resource Unemployment," forthcoming in E. S. Phelps, A. Alchian, C. C. Holt, *et al.*, *Microeconomic Foundations of Employment and Inflation Theory* (New York, 1969).
2. R. W. Clower, "Some Theory of an Ignorant Monopolist," *Econ. J.*, Dec., 1959.
3. D. F. Gordon and A. Hynes, "On the Theory of Price Dynamics," forthcoming in *Microeconomic Foundations of Employment and Inflation Theory, op. cit.*
4. C. C. Holt, "Job Search, Phillips' Wage Relation and Union Evidence: Theory and Evidence," forthcoming in *Microeconomic Foundations of Employment and Inflation Theory, op. cit.*
5. O. Lange, *Price Flexibility and Full Employment* (Bloomington, 1944).
6. A. Leijonhufvud, *Keynesian Economics and the Economics of Keynes* (New York, 1968).
7. R. E. Lucas and L. A. Rapping, "Real Wages, Employment and Inflation," forthcoming in *Microeconomic Foundations of Employment and Inflation Theory, op. cit.*
8. D. T. Mortensen, "A Theory of Wage and Employment Dynamics," forthcoming in *Microeconomic Foundations of Employment and Inflation Theory, op. cit.*
9. E. S. Phelps, "A Note on the Real Wage-

---

[33] This assumes, in Lucas-Rapping fashion, that the money rate of interest at which current quasi-rents can be lent has not fallen, if at all, by as much as (or more than) the expected rate of decline of future prices and wage rates.

[34] Taubman and Wilkinson are verifying and further developing this hypothesis in a forthcoming manuscript

Employment Relation in Zabel's Competitive Inventory Model," *Int. Econ. Rev.* (forthcoming).

10. ———, "Money-Wage Dynamics and Labor-Market Equilibrium," *J.P.E.*, July-Aug., 1968, Part II.
11. E. S. PHELPS AND S. G. WINTER, JR., "Optimal Price Policy under Atomistic Competition," forthcoming in *Microeconomic Foundations of Employment and Inflation Theory, op. cit.*
12. D. H. ROBERTSON, *Money* (Cambridge, 1929).
13. G. J. STIGLER, "Information in the Labor Market," *J.P.E.*, Oct., 1962, Part II.
14. P. TAUBMAN AND M. WILKINSON, "User Cost, Capital Utilization and Investment Theory," Pennsylvania discussion paper, 1968.
15. J. TOBIN, "Money Wage Rates and Employment," in S. E. Harris, ed., *The New Economics* (New York, 1952).
16. S. WEINTRAUB, *Price Theory* (New York, 1949).
17. E. ZABEL, "A Dynamic Model of the Competitive Firm," *Int. Econ. Rev.*, June, 1967.

DISCUSSION

Frank Brechling: These three papers contain important and fundamental contributions to the theory of the Phillips curve. Holt's and Phelps's papers deal with the micro-foundations of the Phillips curve while Archibald addresses himself to the specific problem of the impact of changes in the dispersion of sectoral unemployment rates on the position of the Phillips curve. An adequate evaluation of these three contributions is unfortunately impossible in the limited available time and, hence, I shall confine my comments to a number of specific points.

Following a long tradition in macroeconomics, Phelps argues forcefully that the expected rate of wage increase is a determinant of the position of the Phillips curve. There is a unique rate of unemployment (U^*) at which actual wage increases equal expected wage increases. At unemployment levels below U^* there will be ever increasing wage increases (and, hence, accelerating price inflation) and at unemployment levels above U^* there will be decreasing increases in wages (and, hence, ultimately accelerating price deflation). Since most previous analyses of the Phillips curve tended to ignore this expectations effect Phelps's emphasis on it is most welcome. Before accepting it as part of received theory we should, however, consider the following points:

First, the "expectations hypothesis" implies clockwise loops around the fitted Phillips relationship. In other words, wage increases should be comparatively low in upswings (because expected wage increases are low) and comparatively high in downswings. In fact, however, the loops which have been observed in both Britain and the U.S. are virtually all anticlockwise. Thus an important implication of the expectations hypothesis is not supported by most of the available empirical evidence. To the extent that this hypothesis is valid, its effects must have been overshadowed by some other factors.

Second, there may be factors which offset the upward or downward pressures on wage changes generated by the expectations hypothesis. Suppose for instance, that there is a "learning by working" effect which relates the quality of labor, the knowledge and tastes of workers to the cumulated number of days that they have been in employment, perhaps making due allowance for employment interruptions. A permanent reduction in unemployment then generates ever improving quality of labor, ever increasing knowledge and ever changing tastes which may offset partially the effects of the expectations hypothesis. They may, of course, also reenforce the expectations hypothesis. Before accepting the expectations hypothesis, we should attempt to discover all the long-run effects of fixing the unemployment rate at a particular high or low level.

Third, the expectations hypothesis is much more plausible for hyperinflation than for the comparatively mild inflations which most western countries have experienced in the postwar period. In a world with uncertainty, high costs of information, lagged adjustments, etc., it seems plausible to assume that U^* is not a particular rate of unemployment but rather a range within which expectations are not generated by some dominant factors and, hence, within which the Phillips curve is stable. Any deviation from this range would then trigger off expectations which in turn would lead to explosive inflation or deflation. The existence, location and size of this range should be determined by empirical investigations.

Fourth, it is frequently asserted that the expectations hypothesis destroys the notion of a trade-off between unemployment and wage inflation. There is, however, no reason why slowly accelerating inflation and high output should not be preferable to constant (or zero) inflation and low output. Moreover, a stabilization program might conceivably involve the maintenance of low levels of unemployment interrupted by sharp increases in unemployment to destroy inflationary expectations.

The upshot of the four points is that—although the expectations hypothesis is most important and should be incorporated in all future discussions of the Phillips curve—it is as yet too early to be able to assess its full impact and policy implications. The relevant empirical evidence seems to suggest that the expectations hypothesis is either not strong or overshadowed by offsetting factors. Moreover, it is not very difficult to develop a sketch of a theoretical model in which the expectations effect is either weak or offset by other factors. Further theoretical and empirical work is, therefore, required before the expectations hypothesis can be regarded as part of economic orthodoxy.

In the Holt paper the Phillips curve is derived from the stochastic and dynamic interactions of employers and prospective employees. The higher the level of unemployment, the longer is the average period of unemployment and, hence, the lower is the wage aspiration of the searching unemployed. The declining aspiration levels, in turn, increase the probability that the finally accepted wage is lower than the previously received wage.

According to Holt's model the position of the Phillips curve is primarily determined by the speed and efficiency with which the labor market brings together two partners to a wage bargain. Thus if the probability of placement of an unemployed rises or if the rate of labor turnover is reduced, then total unemployment can be reduced, and, consequently, the Phillips curve shifts to the left. Thus Holt's policy recommendations are aimed at raising the probability of placement and at reducing turnover rates. There are two small points which arise in this connection.

First, a reduction in labor turnover clearly reduces unemployment but it may also increase inefficiencies in the allocation of labor. Thus the benefit derived from a leftward shift of the Phillips curve may have to be considered in conjunction with the cost of increased allocative inefficiency. While Holt recognizes this point, he does not pay much attention to it. More theoretical and empirical work is clearly needed.

Second, the interactions and interdependences in Holt's policy recommendations are quite complex. Suppose some policy measure raises the probability of placement. By itself this would lead to a fall in unemployment. It would, however, also raise turnover rates and, thereby, raise unemployment. Interactions of this kind make it extremely difficult to develop simple policy recommendations. We need to know more about the structure of the interactions as well as the magnitude of the relevant parameters.

Archibald has ingeniously and successfully dealt with an aggregation problem without actually fitting disaggregated relationships. His method consists of introducing second and third moments of the sectoral distributions of unemployment as independent variables into the aggregate relationship. The hypothesis that a rise in the dispersion in sectoral unemployment rates raises the aggregate Phillips curve was first put forward by Lipsey, and Archibald has elaborated the necessary and sufficient conditions for this hypothesis to be correct. I have two minor comments.

First, Archibald's evidence seems to be consistent with the notion of wage leadership, according to which the sectors with the largest excess demands determine wage rate changes for the other sectors. This notion is supported by the negative sign of the regression coefficient of γ (a measure of skewness). If wage leadership is an important phenomenon, then economic policy ought to be aimed at reducing the pressure of excess demand in the leading sectors. A general depression of demand may then be an inefficient policy tool.

Second, the quantitative implications of Archibald's results are quite striking. The mean σ_u^2 for the U.S. sample is about 1.0. The coefficients in Table 6 suggest that if σ_u^2 were to be reduced to zero, then the rate of change in money wages would decline by about 1.7 or 1.8 percentage points per quarter. This figure is unexpectedly large. It suggests that the inflationary pressure may be reduced markedly by eliminating regional dispersions in unemployment.

Archibald has opened up an important and interesting area of research. His results are extremely encouraging and they are likely to lead to significant insights in the discussion of deficient demand and structural unemployment and have important implications for economic policy.

Otto Eckstein: The traditional theory of the rational individual has certainly proved to be the most beloved tool of the economist. Our culture has as its philosophical cornerstone, "Theory should be derived from traditional microeconomics." Derivation of a hypothesis from the classic micro principle is tantamount to empirical testing, or at least the testing can be safely deferred. There is a lack of purity in other approaches, a sinfulness that the young graduate student is well advised to avoid.

The human being is an individual and he does calculate to advance his welfare. The theory is therefore fruitful in predicting his actions and in suggesting aggregate hypotheses. The papers by Holt and Phelps are fine examples of the power of the theory. Holt's analytic development of Stigler's concept of search unemployment and his probabilistic model of wage offers and aspirations yield interesting, Phillips curve-like, conclusions. The concepts provide a more coherent foundation for relating wages to labor market conditions. Many of the conclusions have long been accepted and we can be grateful for a more elegant derivation than has been previously available.

Holt's theory of wages and employment rests on a sophisticated model of frictional unemployment. There is no Keynesian involuntary unemployment, nor structural unemployment. Wages are wholly determined by conditions in the labor market, without any influence from product markets. Institutional studies of labor economics and most econometric studies of wages have found conditions in the product market to be of pertinence for wage movements. I am not familiar with any empirical test which claims to show that wages are set only by labor market conditions. The absence of the link between wages and product markets removes industrial market power as a pertinent, inflationary factor.

The doctrine also leaves little role for unions. To be sure, in Professor Holt's skillful hands there is a fine theoretical derivation of the relative wages of union and nonunion workers over the cycle. But the central mechanism is wholly individual. The formal theory of groups and unions has not proved as tractable mathematically or as fruitful in suggesting hypotheses as the classical theory of the individual. Nonetheless, there is a large body of knowledge about the dynamics of union wage claims and about collective bargaining which the present theory just shrugs aside. The unemployment-search phenomenon would be asked to carry rather too much if one were to attempt to explain wages reached by nationwide bargaining in highly organized and concentrated industries by it and it alone. Stigler's evidence pertained mainly to professionals. And the recent theorizing may be colored too much by introspection based on the peculiar market for college professors.

Finally, the central employment problem of our society today is the disparity of employment opportunities among blacks and whites, among skilled and unskilled, among young and experienced. The reader will look in vain for guidance on these matters. We are in danger of devising a labor market theory which is as remote from the central employment problem of our times as the classical theory was in the 1930's. Once again, search unemployment makes all unemployment voluntary, and its solution lies in the realm of lowering real wages and improving the efficiency of the market.

The paper by Phelps surveys the current classical revival of microeconomics. He almost persuades us that a new theory is emerging, and his exposition is extraordinarily skillful and insightful. A more complete survey would produce further recent examples of the new neoclassics. If the survey were extended in time, reporting on earlier disequilibrium theories of wages, employment and prices, one would see rather more continuity and less newness.

The "new" neoclassics of wages and inflation has one aspect which can easily set us on some wrong roads. This is the emphasis on "real wages" as opposed to money wages. As Professor Phelps says so eloquently, the new theory "makes no appeal to faulty perceptions [money illusion] . . .," and in the writings of other well-known classical writers, it is only a short and barely perceptible step from the general adaptive expectations model to the unitary coefficient case where all decisions are made in real terms, where consumers and businesses have a full and precise knowledge of the future of prices and fully and instantly deflate all magnitudes by the price indexes. Once that leap is taken, the Phillips curve disappears altogether, the society faces no trade-offs between unmployment and price stability, the "natural" unemployment rate becomes inescapable, and a ruthless macro policy of orthodox deflation is easily recommended.

At the risk of muddling the semantic situation, let me argue that it is money wages which are real and "real" wages which are fictitious. As Keynes contended long ago, people pay, receive, bargain, and strike over money wages. The price indexes used to compute real wages, whether industry-own or consumer, are the creation of statisticians, containing unknown biases, never behaving like a predictable trend. Every serious student of the wage problem of the last ten years has agreed that consumer prices have an impact on wage claims, and that the Phillips curve at any moment in time is affected by the previous price history. But that is saying very much less than the now fashionable view that the discounting of inflation is quick, complete, and correct, that wages and labor supply can be better understood by focusing only on "real" wages, that the "expectations" hypothesis is an established empirical reality.

Somehow, early on, the term "money illusion" crept into our economists' dictionary. Decisions in money terms became "irrational" whereas decisions deflated by various price indexes became "rational." But what is the nature of the illusion? The definition may be clear in the perfect information case, where price change is known, but that is surely not the relevant case. Who among us knew that consumer prices would rise by 4¾ percent in 1968? The forecasts among the experts for 1969 range from less than 3 to 5 percent.

Professor Archibald asks "whether or not there is a Phillips curve" and seeks an answer by statistical tests of the Phillips-Lipsey hypothesis with and without consumer prices. It is a paper of the highest craftsmanship, and in some respects, his paper is an answer to the "new" neoclassics.

In a way, the question of the existence of the Phillips curve itself has come to mean different things to people. The really simple-minded Phillips curve cannot be attributed to Phillips, who included consumer prices beyond a threshold rate of advance. Lipsey rather chipped away at the specific role assigned to prices, but still left them a general, though weak, role.

The advocates of profits and value productivity as wage determinants denied the Phillips curve on the ground that their factors matter, too, and that the Phillips curve is really a family of curves. The "new" view emphasizes the role of price expectations rather than price history and generates a family of Phillips curves on this new factor. While no doubt a point of substance, the empirical validity or importance remains unknown. Among empirical workers, the dispute over the existence of a Phillips curve thus is a limited one —a disagreement about the variables that matter besides unemployment, and a subsidiary question about the value of the coefficient on consumer prices, the time profile of the effect, and the nature of the expectations process.

The substantive scientific issues are quite clear and the logical differences small. It is the policy implications that are divisive. Those who stress the Phillips curve rather than its qualifications want to move the curve, through structural policies and/or exhortation. Those who stress the qualifications—or deny the trade-off altogether—believe the economic structure needs no remedy. They would solve the inflation problem entirely by limiting the growth of demand and applying a dose of deflation so severe as to destroy the (unmeasured) psychology of inflationary expectations. An eclectic might argue that in rapid inflation, such as the present, policy must concern itself first with limiting total demand and the restoration of more moderate price expectations. But the structural problems demand attention too. The inflation problem of our day is the result of the action of excessive demand on an imperfect economic structure.

RICHARD A. LESTER: Although, in general, I agree with Professor Holt's policy positions, I question his assumptions with respect to the dominant dynamics of wage changes in the short run, particularly for firms with 100 or more employees, which have over three-fifths of all employees in the country. (Firms with 1,000 or more employees account for over two-fifths of all employees.)

Studies of the wage- and benefit-setting process in medium-sized and large firms and studies of labor mobility have convinced many labor economists that interfirm labor mobility and changes in individual workers' supply prices are not the most important factors in the determination of wage structures, wage movements, and company benefit costs in the short run. More important for short-run changes in the level of wages and employee benefits are, apparently, such factors as collective bargaining, the threat of unionization, and discontent and demonstration on the job, without labor mobility in the form of job shifts. In other words, group forces and collective action are likely to have more influence than individual calculation and independent personal action in determining the size and timing of wage-benefit changes and the structure and composition of wage-benefit packages in the short run, say from year to year or over a two-year period.

The figure of only one-third of all employees covered by collective agreements is not a measure of the wage-benefit influences of collective action and collective agreements. Pattern-setting negotiations tend to determine the size and composition of wage-benefit increases in a particular year in many nonunion as well as unionized firms, and most unionized workers are covered by negotiated agreements that run for two or three years. Under them, the wage-benefit structure and increases are predetermined for that period largely independent of short-run changes in market conditions. Such institutional factors cannot be ignored.

Not only are nonunion firms of any size likely to be influenced by the results of collective bargaining in unionized firms in the industry and the locality, but such nonunion firms are apt to have wage structures (occupational differentials) fixed by systems of job evaluation, and to give wage increases in uniform cents per hour or a percentage for all occupations. Moreover, unionized workers, on the average, receive more frequent general wage increases and more frequent revisions in benefit programs than do nonunion employees. During the period from 1959 to 1966, for instance, the rate of introduction or revision of benefits in union factories was about double that in nonunion plants (*Monthly Labor Review*, Feb., 1968, p. 48).

It is such institutional and bargaining factors that help to explain why, for manufacturing during the 1951–67 period, the change from one year to the next in the rate of unemployment and in the rate of increase in negotiated wage gains was in the same direction (both moving up or down together) in almost half of the years, as indicated in my article in the May, 1968, *Review of Economics and Statistics*. Such short-run unemployment-wage behavior was, of course, contrary to Phillips curve expectations, but the longer-run wage-unemployment behavior was much closer to Phillips curve predictions. And there can be little doubt that, in recent years, wage movements and wage structures have been greatly influenced in the short run by threats of collective action and by negotiations in such industries as building construction, trucking, coal, steel, autos, and even in public education, police protection, and garbage collection in large cities.

Recognition of the significant short-run role of collective factors is not to deny the influence of market forces on wage scales, wage movements, and benefit programs in small enterprises in the short run and in medium-sized and even in large firms in the long run. Nor does the increasing significance of collective elements deny that some measure of adjustment to market forces may occur in companies during the life of two- and three-year agreements by such means as raising and lowering hiring standards, hiring in at a point higher up in the wage structure during times of labor stringency, the tightening or loosening of production standards with changes in market conditions, etc. It is maintained, however, that, in terms of the dynamics of year-to-year changes in wage-benefit packages, such on-the-job adjustments are of secondary importance, and that, for the problem of the wage-price spiral in the short run, collective factors (demonstrations, strike threats, strikes, negotiated agreements, compliance with "orbits of coercive comparison," etc.) are likely to be more powerful influences on wage-benefit structures and the size of wage-benefit increases than are individual worker search, single workers' aspirations and acceptance wages, individual employers' aspirations for output per wage dollar, etc.

Professor Phelps generally assumes the same sort of labor market dynamics as Holt. Individual wage rates (or is it the company's whole compensation schedule?) are supposed to move up and down with the volume of "search unemployment" and relationships between searchers' "acceptance wages" and "actual wage rates." (Should "money wage rates" be interpreted to mean the rate for a particular hiring-in job or the recruit's potential earnings over a work life in that firm?) The idea that idle suppliers of labor, by individually raising their wage expectations or standards, generate a succession of wage increases (personal or occupational rates or the establishment's whole structure of compensation including benefits?) is unclear both in conception and in assumed operation with respect to established firms of any size. Certainly such notions are subject to serious question on the basis of labor market studies and studies of company and industry wage behavior. Apparently it is not generally understood that the volume of a firm's job applications during a week, month, or year is not closely related to its location in the community's wage structure and that job applications, even at the highest paying firms, tend to dry up if those firms are not hiring or are laying off employees.

For short-run policy questions, the microeconomics of labor must recognize important differences in the process of employment and in the determination of worker compensation in various industries. The processes of hiring and advancment are rather dissimilar in such industries as building construction, coal, railroads, auto assembly, postal service, police protection, and secretarial employment. Also, among those industries significant contrasts exist in wage structures, wage determination, and wage movements. For example, hiring for construction of commercial and public buildings in cities is under tight union control, and union wage movements in construction in recent years have had their own peculiar pattern apart from manufacturing, coal, railroads, etc. Indeed, the sharp rises in negotiated building rates in 1967 and 1968—much more rapidly than for manufacturing and other parts of the economy and than could be explained by relative unemployment in building—has brought home the extent to which construction under union control is a law unto itself.

Although in his analysis of the dynamics of wage-benefit developments the theorist may be reluctant to descend to the industry or large-firm level, it does not make too much sense to apply the assumed processes of adjustment in the Holt model to short-run wage changes in large firms or to changes in benefit programs, which are applied uniformly (generally on a company-wide basis) and are not subject to personal negotiation, modification, or exception.

Moreover, it does not make much sense, for example, to apply to the following parts of the economy notions of individual workers raising wage scales by means of the level of their "acceptance wages" and employers adjusting individual wage rates according to such acceptance wages: the organized part of the construction industry, with its own hiring halls and adherence to standard rates; the basic steel industry, which has its own formal wage structure jointly administered; the large auto, rubber, oil, electrical equipment, aircraft, and chemical companies, which have carefully worked out wage structures and whose wage-benefit increases tend to conform to a standard pattern; and the coal, railroad, maritime, and over-the-road trucking industries where the unions tend to dominate wage-benefit structures and movements.

On the other hand, it may make sense to apply such a model of the dynamics of wage determination in areas of the economy that lack formal wage structures and are unorganized, such as local trucking, secretarial and clerical occupations (especially in small enterprises), small-scale metal fabrication operations that are not part of an industry orbit, agriculture, small-scale service and retail operations, domestic service, etc.

With respect to bargained levels of wage-benefit compensation, there is an extensive literature showing the need for a framework of analysis that includes product and labor market forces, union political pressures, and the bargaining power of the parties. Harold Levinson has summarized that literature and applied such an analytical framework to experience in five industries in his recent book on *Determining Forces in Collective Wage Bargaining*. In any analytical examination of short-run, wage-benefit behavior, that body of literature should not be disregarded.

In addition, there is need for further refinement and development of the microeconomics of employment and worker compensation as applied to medium-sized and large firms, whether unionized or not. Such enterprises constitute dominant elements on the demand side of the market for most occupational skills and for unskilled labor to be trained on the job.

Some labor economists have pointed out that the "wage market" (the process of wage-rate change) and the "job market" (the process of individuals moving among jobs) may have rather independent existences. Especially in highly unionized industries and in large firms, the influence of labor mobility on wage-benefit structures and levels may be rather tenuous, or only significant beyond a certain range of tolerance. In both model building and public policy, differences between the factors determining worker compensation and the factors determining employment allocation must be assessed and taken into account. That is a clear-cut lesson from both research and experience.

THE SUPPLY FUNCTION IN AGRICULTURE REVISITED

THE REPRESENTATIVE FARM APPROACH TO ESTIMATION OF SUPPLY RESPONSE

By Jerry A. Sharples
United States Department of Agriculture

In this paper I discuss the role that representative farms have played in the estimation of aggregate supply response, and I speculate on the future role of representative farms in aggregate analyses. I do not discuss the valuable contribution or the problems associated with the use of representative farms in firm analyses per se.

Many empirical techniques can be used to estimate aggregate supply response. They include regression, spatial equilibrium, mathematical programming, and less formal techniques. However, the representative farm approach has been by far the most popular research approach used for both firm and aggregate supply response analysis during the 1960's. True, other techniques were used, but no model in the history of production economics has received as much widespread research support in the form of research funds and professional manpower. I am referring to the regional adjustment studies and the associated state studies. In all these studies representative farms were used to estimate aggregate supply response. Since most of the regional adjustment studies have been completed, it is time to evaluate how well this approach to supply response has worked, and to speculate about where we should go from here in aggregate supply analysis.

Representative Farms and Aggregate Supply Response: Last Ten Years

The history of supply response research and the role played by representative farms are well documented.[1] I would like to break into the historical chain of events at about the late 1950's. At that time there was discussion about a supply response model that was based on the aggregation of representative farm results—the model that later became the foundation of the regional adjustment study research. This model had five major components. These were, following Barker and Stanton [1, p. 701]: (1) stratify all farms within a region into homogeneous groups, (2) define a representative farm for each stratum, (3) derive supply functions for each farm, (4) aggregate the supply functions, and (5) remove the model's simplifying assumptions and adjust the results accordingly in order to make predictions or prescriptions. I can find no label in the literature for this model, so I will call it a "representative farm aggregation" model (RFA model).[2]

The RFA model had an intuitive appeal. First, it seemed to provide the missing research link between the firm and the industry—the micro and the aggregate. The impact of a change in a variable could be traced from the farm where the production decisions are made, up to the industry where policy decisions are made. A model that could predict aggregate response via the micro route appeared to be a powerful tool.

---

[1] See Barker and Stanton [1], Nerlove and Bachman [7], the book on agricultural supply functions edited by Heady and others [6], and the proceedings papers and discussions of the use of representative farms in agricultural economics, at the summer, 1963 A.F.E.A. meetings [3] [5] [9].

[2] I use "FRA model" to refer to a model where (*a*) the unit of analysis is obviously the firm, i.e., the restrictions and alternative activities in the model reflect a firm's restrictions and alternatives, and (*b*) the primary objective is the estimation of aggregate supply response.

Second, the normative characteristics of the RFA model would allow its use for studying conditions that had not occurred in the past. This model could overcome the great limitation of regression analysis.

The RFA model was not new in the late 1950's, but two research innovations caused renewed interest in the model: linear programming and electronic computers. It would have been a formidable task to derive supply functions for many representative farms by hand using conventional budgeting techniques. But LP and computers overcame this problem. Thus, the profession got excited about the potential of the RFA model as a research tool. It appeared that even though many theoretical and technical problems were anticipated, we were surely at the dawn of a new era in understanding aggregate supply response.

In 1958 the first of six regional adjustment studies—the Lake States dairy study—was initiated. Over the next five years similar adjustment studies were started in the Northeast, South, Corn Belt, Great Plains, and West. The RFA model was the foundation for each of the six regional studies. The Lake States dairy study was completed in 1963.[3] The last of the regional adjustment studies will probably be completed by 1969.

These studies were cooperative ventures by the state agricultural experiment stations and the Farm Production Economics Division of the Economic Research Service, U.S. Department of Agriculture. The primary objectives of all the regional adjustment studies were (1) to estimate adjustments needed in resource allocation on farms to maximize profits, (2) to estimate regional supply response, and (3) to determine an equilibrium supply-demand condition for a region and then for individual farms. The U.S. Department of Agriculture had special interest in short- and intermediate-run regional analyses and the state experiment stations had special interest in the farm, substate and state analyses.

The great expectations that the profession had for the RFA model as used in the regional adjustment studies is shown by the large expenditure of research resources we were willing to invest. A rough measure of this research support is the number of published reports and the number of M.S. and Ph.D. theses connected to the regional adjustment studies. For example, from the Southern (S-42) and Corn Belt (NC-54) adjustment studies combined came over eighty research publications, forty M.S. theses, and twenty Ph.D. theses. Many of the theses are included in the count of publications. Additional publications and theses came—and are still coming—from the other adjustment studies but a tally of them is not available. The profession will, for many years, be feeling the impact of having a large number of their production economists trained via the regional adjustment studies.

Much has been gained from the regional adjustment studies. First, but possibly not foremost, were the representative farm adjustment results and the regional results. Second, a data base (budgets, representative farm data, and other quantitive data) was established that is currently being utilized and updated in other research projects. Third, these studies provided graduate training, as noted above, and technical training as well. Many in the profession got their first taste of mathematical programming and electronic data processing via the regional adjustment studies.

---

[3] The Lake States results are summarized in [12]. Examples of other regional results are [4] [8] and [10].

And fourth, we learned that the problems connected with the RFA model were more devastating than we had first imagined. The aggregate regional analysis—the justification for having regional projects—was disappointing. I think that the disappointment, however, was caused more by the original great expectations than by actual performance.

Problems with the RFA model

Several writers have dwelled on the problems associated with the RFA model as used in the regional adjustment studies.[4] The problems most difficult to overcome can be grouped into five categories: (1) interdependence or externalities, (2) change in farm size, (3) unrealistic firm-level assumptions, (4) the selection of representative farms, and (5) mechanical problems.

Interdependence or Externalities. In the RFA model used in the regional adjustment studies, aggregate supply response is obtained by the summation of weighted representative farm supply functions. The individual farm's input costs and physical transformation rates are assumed to remain constant at all levels of aggregate production. But we know that the aggregate level of production has a major impact on input prices which in turn affects each farm's supply function. The impact is greater as larger regions are analyzed. If we were studying supply response of corn and our analysis were limited to Michigan, a 20 percent change in corn production in Michigan would not have a major impact on input prices, but if the Corn Belt were our area of analysis, a change in corn production of 20 percent would alter input prices considerably.

Attempts are now being made to overcome this problem by putting all representative farms into one programming model. The major inputs can then be limited on a regional or subregional level, or they can enter the matrix as a regional stepped supply function. The representative farms could then compete for inputs such as beef feeders, feed, and labor. However, this approach creates about as many problems as it solves. How do you define the regional or subregional supply of, say, beef feeders? Are you willing to live with a solution that maximizes area profits instead of firm profits? Also, the size of the computational problem becomes a major factor in limiting the number of representative farms.

Change in Farm Size. A major factor associated with aggregate supply response is the changing size of the farm. In the RFA model it is very difficult to evaluate the aggregate impact of allowing representative farms to buy or rent additional land. There is no way to insure that purchased land equals sold land or that land rented in equals land rented out. In an attempt at overcoming this problem, several studies projected the distribution of farms by size to the target date of the study. The representative farms could then be aggregated using projected aggregation coefficients. But these attempts do not show which farms would change size or how the purchase or rental of additional land would compete with alternative farm uses of capital.

Unrealistic Firm-level Assumptions. Many of the assumptions of the RFA model can be grouped under this problem heading. These include: farmers are profit maximizers, farmers have no enterprise preferences, farmers have perfect knowledge of their alternatives, uncertainty does not enter into the choice criteria, and so on. These assumptions pertain to the individual representative farms but they have profound aggregate implications. Alternative assumptions could be examined. For example, a model that minimizes the cost of obtaining an income goal may be appropriate. Or it could be assumed that only those enterprises currently on a farm would be included in the model; i.e., livestock enterprises would not be alternatives on a representative cash grain farm.

---

[4] For example, see Barker and Stanton [1], Sharples, Miller and Day [11], or Brokken [2].

Selection of Representative Farms. The selection of representative farms and the associated problem of aggregation error have probably been the subject of the most thorough and diligent research of any of the problems listed above. These studies have shown that the number of representative farms needed to eliminate aggregation error approximates a geometric function of the complexity of the model. As more restraints and activities are defined, many more representative farms are needed. Thus one can only strive to minimize aggregation error, not eliminate it. These studies have also shown that the traditional methods of stratifying farms—by farm size, soil type, or farm type—may lead to a set of representative farms all of which have about the same optimum organization. Recognition of differences in institutional restrictions, motivations, preferences, ability, and other human factors needs to be built into the specification and stratification schemes.

Mechanical Problems. A few words need to be addressed to the purely mechanical problems associated with a model that is used to estimate aggregate supply response by summing representative farm supply functions. Only a dedicated Ph.D. candidate would have the incentive to assemble all the budgets and restraints necessary to build a reasonably rigorous model. And once the budgets and restraints are defined, the matrices are complete, and the computing is underway, a second mechanical problem arises: how will one analyze all the computer output. It takes great skill to keep from being buried in computer print-out. These technical problems cause a researcher to think twice about making any improvements in the RFA type of model that would lead to more representative farms or more activities and restrictions. Also, as one tries to build more realism into the model, he risks making the model so complicated that he cannot readily trace the logical connectives between a change in an instrument variable and a resulting change in production. A model's results are of little value if they are not accompanied by a statement explaining the causes of changes.

Future Role of Representative Farms in Aggregate Supply Response Research

This brings us up to date. We know that the RFA model as used in the regional adjustment studies has serious limitations—as do all models—and that currently there is some research endeavor to better understand the limitations. But where do we go from here in aggregate supply response analysis? What role should representative farms play in future supply response research? In order to address these questions, we need to differentiate between short-run and long-run supply response.

I do not expect the RFA model to contribute much more to expanding our knowledge of intermediate and long-run aggregate supply response. Two considerations led me to this conclusion. First, the coefficients in the RFA model reflect an assumed level of technology, an assumed institutional structure, and an assumed size of farm. Given these assumptions, the RFA model focuses on profit maximization. This model has empirical relevance. But the most dynamic ingredient affecting supply response is the changing structure of agriculture—changing technology and the changing institutional framework within which agriculture functions. Representative farm models look at only one point in the whole spectrum of structural change. A more appropriate model would be one that focuses more on structural changes and less on profit maximization and movements toward equilibrium.

Second, the representative farm becomes much more difficult to define as economic time (and calendar time) becomes longer. Production coefficients, resource ratios, the objective function, and other farm characteristics are usually based on existing farms. Adjustments can be made in the parameters for some assumed increase in efficiency. But the present structure of agriculture pervades the data. When current data are used for describing a representative farm, a structure is assumed for the credit market, the input markets, the product market, government policy, and so on. If one were to define representative farms for some point in future time, he would have to speculate on the whole structure of agriculture at that point in time before he could develop his coefficients. Because of the dynamics of agriculture, the longer the economic time period, the harder it is to specify characteristics of individual farms.

I must admit that I have no better model to put forth. Besides, that is outside the realm of my assigned topic. However, I think that we should put relatively more effort into intermediate- and long-run aggregate supply response research using aggregate data and aggregate relationships.

A conditionally normative representative farm approach to aggregate supply response is still intuitively appealing for certain types of short-run questions. For example, we may wish to analyze a one-year response of farmers to a change in government policy relating to diversion of wheat allotment acres. What aggregate changes would take place in wheat production, production of other crops, resource use, and farm income? In order to answer these questions adequately one must look at aggregate data and aggregate relationships, but one can get valuable insights into the nature and magnitude of the aggregate relationships by examining the effect of the policy change on several well-chosen representative farms. Also in the short run (both in terms of short-run economic time and in terms of one to three years, calendar time) technological and institutional change are much more predictable than in the longer run. Thus, the representative farm identification problem is less difficult to handle than in the long-run analysis. We can estimate production coefficients, resource ratios, and how many farms there are with various characteristics assuming the current structure of agriculture.

Up to now there has been virtually no use of representative farms for analyzing short-run aggregate supply response. The reasons are twofold: (1) The primary objective of state or regional supply response studies has been to analyze state or regional adjustment implied under normative equilibrium conditions. This objective necessarily implied a longer-run model. And (2) it took a long time to collect data, define representative farms, compute, aggregate, and analyze the results. There was always the fear of finishing the study after the study projection period had passed. So projection dates had to be at least five years beyond the starting date of the study, or one might end up projecting to a time period that had already passed.

In order to use representative farms as an aid in short-run aggregate supply analysis, we will need to change our concepts of the role of representative farms in aggregate analyses. First, new ways will need to be developed for linking firm response to aggregate response. A simple horizontal summation of weighted representative farm supply functions has not proven satisfactory. Second, the model will have to be simpler—not more complex—than we envisioned in the past. Data handling, computing, and analysis of results will have to be streamlined, so that the research can be done quickly. For short-run analysis, timeliness of results is vital. In order to be timely, the model must contain relatively few representative farms—more in the nature of 5 to 10 than 100–400 as in the regional adjustment studies. Experience with the adjustment studies showed that the major economic relationships—those relationships that the researchers were trying to isolate—did not differ greatly over the 100–400 representative farms. Most of these relationships could have been isolated by some parametric resource and price programming on fewer representative farms.

As an example of how this model might work, consider the following: suppose we wanted to analyze how an increase in price support rates in a government feed grain program would affect Iowa farmers. In order to adequately answer this question we would need to estimate changes in production, resource use and income at the state level and also give some description of which farmers would be affected. There are considerable data available at the state level to give a rough indication of the magnitude of change. Historical data series at the state level could give some indication of farmers' willingness to adjust their cropping patterns in the past to changes in instrument variables. Several representative farms in Iowa could also be analyzed to see how increased corn price supports would affect optimum production, resource allocation, and income at the firm level. Information gained at the firm level could be used as an aid in adjusting the aggregate estimates.

I am admittedly vague and sketchy about the linkage between the micro case-farm type of analysis and the aggregate analysis. All I wish to point out is that one can gain insights from a limited firm analysis that is useful for understanding short-run supply response at the aggregate level. We need not ignore this micro information just because it cannot be plugged into a neat mathematical aggregation formula. This type of short-run analysis is currently being tested by the Aggregate Production Analysis Team in the Farm Production Economics Division of ERS.

Conclusions

Barker and Stanton observed in 1965 that "we are still a long way from developing usable estimates of national or regional supply from individual farm data. Is it realistic to hope to use individual farm data in estimating agricultural supply? In the mid-60's the answers are not yet clear. . . . The logic of building from the bottom up still seems to outweigh all the practical problems of data collection, programming and aggregation" [1, p. 712].

Over three years have passed since this observation was made. Much research has gone into the "practical problems." Now we better understand these problems, but they are no less formidable. The evidence is mounting that the RFA model may not play a major role in future research of intermediate and long-run aggregate supply response. The RFA model did not turn out to provide the big research breakthrough that we had originally hoped.

The representative farm can play a potentially important role in short-run aggregate supply response. Knowledge of the potential economic impact of a change in an instrument variable on a farmer's income and organization is very valuable. However, the linkage between the firm and the aggregate may by necessity be an informal one.

Nerlove and Bachman in 1960 concluded their review of supply response research with the observation that "there is a need for a combination of empirical approaches to the supply problem at all levels of aggregation between firm and industry" [7, p. 533]. During the 1960's we have been preoccupied with one empirical approach—the RFA model. Supply response is better understood by more professionals because of the extensive use of the RFA model. But the profession paid a high price by not having a more diversified approach to supply response research.

A shift in emphasis is overdue. In terms of better understanding and more accurately quantifying aggregate supply response, I think the payoff will be small from future use of the RFA model. We do not need another round of regional adjustment studies using the RFA model. What we do need is to further explore less cumbersome models that identify and better utilize stable aggregate relationships. We also need to further explore the heart of intermediate and long-run aggregate supply response: technological and institutional change.

REFERENCES

1. RANDOLPF BARKER AND BERNARD F. STANTON, "Estimation and Aggregation of Firm Supply Functions," *J. of Farm Econ.*, Aug., 1965, pp. 701–12.
2. R. F. BROKKEN, "Review of Linear Programming Analyses of Regional Production Adjustments and Supply Response," presented at the W-54 GP-5 Joint Symposium in Fort Collins, Colorado, Aug. 6–8, 1968, to be published as part of the symposium proceedings.
3. HAROLD O. CARTER, "Representative Farms—Guides for Decision Making?" *J. of Farm Econ.*, Dec., 1963, pp. 1448–55.
4. DALE COLYER AND GEORGE D. IRWIN, *Beef, Pork, and Feed Grains in the Corn Belt*, Mo. Agr. Exp. Sta. Res. Bul. 921 and North Central Reg. Res. Pub. No. 178, Aug., 1967.
5. LEE M. DAY, "Use of Representative Farms in Studies of Interregional Competition and Production Response," *J. of Farm Econ.*, Dec., 1963, pp. 1438–45.
6. EARL O. HEADY, *et al.*, *Agricultural Supply Functions* (Iowa State Univ. Press, 1961).
7. MARC NERLOVE AND KENNETH L. BACHMAN, "The Analysis of Changes in Agricultural Supply: Problems and Approaches," *J. of Farm Econ.*, Aug., 1960, pp. 531–54.
8. NORTHEAST DAIRY ADJUSTMENTS COMMITTEE, *Dairy Adjustments in the Northeast*, New Hampshire Agr. Exp. Sta. Bul. 498, June, 1968.
9. JAMES S. PLAXICO AND LUTHER G. TWEETEN, "Representative Farms for Policy and Projection Research," *J. of Farm Econ.*, Dec., 1963, pp. 1458–65.
10. S-42 DRAFTING COMMITTEE, *Cotton: Supply, Demand, and Farm Resource Use*, Ark. Agr. Exp. Sta. Southern Coop. Bul. 110, Nov., 1966.
11. JERRY A. SHARPLES, THOMAS A. MILLER AND LEE M. DAY, *Evaluation of a Firm Model in Estimating Aggregate Supply Response*, Iowa Agr. Exp. Sta. Res. Bul. 558 and North Central Regional Research Publication No. 179, Jan., 1968.
12. W. B. SUNDQUIST, *et al.*, *Equilibrium Analysis of Income-Improving Adjustments on Farms in the Lake States Dairy Region, 1965*, Minn. Agr. Exp. Sta. Tech. Bul. 246, Oct., 1963.

POSITIVISTIC MEASURES OF AGGREGATE SUPPLY ELASTICITIES: SOME NEW APPROACHES

By LUTHER G. TWEETEN *and* C. LEROY QUANCE
Oklahoma State University

Supply elasticities indicate the speed and magnitude of output adjustments in response to changes in product price. The elasticity parameter for aggregate farm output is especially important for public policy, because it measures the ability of the farming industry to adjust production to changing economic conditions continually confronting it in a dynamic economy.

Farmers have considerable latitude to substitute one commodity for another in production over a long period. Eventually, this should lead to adjustments among commodities until comparable resources are earning similar rates of return in production of each commodity. Public policies concerned with the earnings of all farm resources and total farm income must consider the aggregate response of farm output to changing economic conditions. The aggregate response of output to price depends on total resource adjustments in agriculture. Because farm resources are adjusted much more easily among farm commodities than between farm and nonfarm commodities, it follows that the aggregate supply response, which tends to determine total resource earnings in agriculture, is less than the supply response for individual commodities.

Questions we attempt to answer in this paper include: What is the magnitude of the elasticity of U.S. farm output with respect to price in the long and short run? Is the elasticity increasing over time as the ratio of purchased to nonpurchased inputs increase? And are farmers less responsive to price decreases than to price increases as fixed asset theory implies? Three positivistic approaches are used to estimate the aggregate supply elasticity. The first is a direct least squares estimate of the aggregate supply function. One advantage of this approach is that it is straightforward and does not entail problems of aggregating estimated supply parameters with associated problems of substitution and interaction among the components. But this method has the disadvantage of high correlation among the time series and difficulty in separating the influence on output of technology and durable resources—hence, it is difficult to determine the long-run supply elasticity.

Two additional disaggregate approaches which circumvent at least some problems of intercorrelation are also used to estimate the supply elasticity. One involves direct least squares estimates of the separate yield and basic production unit components of crops and livestock. These components are then aggregated to form the elasticity of total supply. The second disaggregate approach is to compute the elasticity of supply from production elasticities and demand elasticities for eight farm inputs. These latter two approaches add to our knowledge of component sources of aggregate supply response.

Direct Least Squares Estimates of the Aggregate Supply Function

Estimates of the aggregate U.S. supply function of farm products presented in Table 1 were obtained by regressing output on relative prices, productive assets, and technology as defined below:

O^t: The dependent variable, output, measured as the index of aggregate farm output (1957–59 = 100) [14]. In one equation, lagged output

TABLE 1

SUPPLY FUNCTIONS FOR AGGREGATE FARM OUTPUT, O, ESTIMATED BY LEAST SQUARES WITH ANNUAL U.S. DATA FOR SPECIFIED TIME PERIODS; COEFFICIENTS, STANDARD ERRORS (IN PARENTHESES), AND RELATED STATISTICS

Equation	Time Period	R^2 d'†	Constant	P_R/P_p^{t-1}	S_p^t	T	O^{t-1}
(1)	1921–66*	0.993 .871	−30.70	0.0693 (.0194)	0.2167 (.0965)	1.0449 (.0559)	
(2)	1921–41	.890 .968	−69.23	.05326 (.0327)	.6336 (.4441)	1.1395 (.1007)	
(3)	1948–66	.982 .465	−57.28	.1380 (.0811)	.4901 (.2365)	.9587 (.0862)	
(4)	1921–66	.993 .932	−26.41	*1921–32* .0738 (.0179) *1933–47* .0628 (.0175) *1948–66* .0695 (.0165)	.0969 (.1158)	1.1134 (.0655)	
(5) (Increasing prices)	1921–66	.992 1.1108	−45.54	.1066 (.0267)	.4264 (.1312)	.9425 (.0755)	
(6) (Decreasing prices)	1921–66	.996 .962	−19.00	.0502 (.0173)	−.0002 (.0974)	1.1521 (.0561)	
(7)	1921–66	.993 1.442	−18.48	*Increasing prices* .0554 (.0154) *Decreasing prices* .0538 (.0163)		.9667 (.0642)	.1776 (.0563)

* Observations for the years 1942–47 are omitted.

† The Durbin-Watson autocorrelation statistic d'. The test for autocorrelation is inconclusive in equations (2), (5), and (7); and is significant at the .05 probability level (two-tailed) in the remaining equations. Computing a first order autoregressive scheme was precluded by lack of time.

O^{t-1}, is used as an independent variable.

P_R/P_p^{t-1}: The lagged index of the ratio of prices received by farmers for crops and livestock to prices paid by farmers for items used in production [3] [12]. When P_R/P_p^{t-1} is for a specified period or for years of declining or increasing relative prices, it is the actual observation for the specified years but has a zero value for other years in the period observed (1957–59 = 100).

S_p^t: The beginning of the year stock of productive farm assets, including real estate, machinery, feed, livestock, and cash held for productive purposes; measured in millions of 1957–59 dollars [12] [13].

T: Productivity index; the ratio of farm output to all farm production inputs in the current year (1957–59 = 100) [12] [13]. This productivity variable reflects changes in management and technology as well as fluctuations in productivity induced by weather.

Two approaches were used to test for differences in supply structure over time and between increasing and decreasing price phases. One was to estimate separate equations for component periods and use an F test to determine if the structure was unchanged. Another approach was to divide the price variable into component periods but leave other variables unchanged in the single least squares equation for the entire period.

In equation (1) (Table 1), output is expressed as a function of relative prices, the stock of productive assets, and productivity with observations over the 1921–66 period, excluding 1942–47. The variables account for a high proportion of the variation in output and in all instances, the hypothesis that the respective coefficients are equal to zero is rejected at the 0.05 probability level.

The same specification is repeated in the next two equations with equation (2) estimated with 1921–41 data and equation (3) with 1948–66 data. Standard errors increase markedly compared to those in equation (1) and statistically, only the coefficients of T in the two equations are consistently different from zero. An F value, calculated to test the hypothesis that the respective coefficients in the two equations were equal, was nearly one. Hence, there is no basis to conclude that the structure of supply changed between the postwar and prewar periods.

The limited observations in the equations segmented by time periods poses a problem. To conserve degrees of freedom and to obtain a more precise test of possible changes in the response of output to price, equation (1) was reestimated with the price series divided into three segments. The first price variable contained actual observations for the 1921–32 period, zeros elsewhere. The second contained actual price observations for the 1933–47 period, zeros elsewhere. And the same procedure was used to construct a third variable for price in the 1948–66 period. The resulting specification in equation (4) explained no more of the variance than that in equation (1). The price coefficients were significantly different from zero at the 0.01 level, but a statistical test is hardly needed to conclude that the coefficients of the segmented price variables do not differ significantly. In short, there is little basis from equations (1)–(4) to conclude that the response of output to price has increased over time.

Equations (5), (6), and (7) are included to determine whether the supply structure differs between years of increasing and decreasing prices. The coefficients in equation (5), estimated with twenty-three observations from years of increasing prices in the 1921–66 period, are significantly different from zero at the 0.01 probability level. The coefficients of the price and productive assets variable in equation (6), estimated with twenty-three observations from the years of decreasing prices in the 1921–66 period, are smaller in magnitude and statistical significance than their counterparts in equation (5). An F test was used to evaluate the hypothesis that the structure of supply was unchanged between increasing and decreasing price years. The hypothesis that the respective coefficients in the two equations were equal is rejected at the 0.10 level but not at the 0.05 probability level. Since the probability was small (less than 10 percent) of obtaining an F ratio so large if the structure were identical for increasing and decreasing prices, some support is provided for the role of fixed asset theory in explaining aggregate supply response.

Equation (7) expresses farm output as a function of productivity, lagged output, one price variable for decreasing prices and another price variable for increasing prices. Coefficients of all variables differ significantly from zero at the 0.01 probability level. But coefficients of the segmented price variables are nearly identical, giving no basis for concluding that the response of farm output to price differs for rising and falling prices.

Several variants of the equations in Table 1 were estimated. One was to divide lagged output into segments and leave price and productivity in their original form for the 1921–66 period. Dividing lagged output into two variables associated respectively with increasing and decreasing prices permitted evaluating the hypothesis that the rate of adjustment of output to the two variables differed between price phases. Segmenting lagged output into three variables associated respectively with the 1921–32, 1933–47, and 1948–66 periods as with price in equation (4) provided a test of the hypothesis that the adjustment rates differed among these periods. Similarity of adjustment rates provided no support for either of the two hypotheses.

Most equations in Table 1 were also estimated with data coded into logarithms. The R^2's of the logarithm equations were generally below those for the equations in Table 1. Also, the equations estimated in logarithms provided no support for the hypothesis that the elasticity of supply is increasing over time or that the elasticity of supply is greater for increasing prices than for decreasing prices.

Elasticity of Supply

We now examine supply elasticities estimated from equations in Table 1. The rising proportion of farm inputs comprised of highly productive and flexible purchased inputs including fertilizers, pesticides, and improved feed supplements suggests that the supply elasticity rises over time. On the other hand, farming is becoming more specialized. If an increasing proportion of farm output is produced by highly specialized broiler, pork, beef, or feed producers who continue to operate at full capacity over a considerable range of price ratios, supply elasticity will fall. Furthermore, the elasticity of supply $(\delta O/\delta P)\ (P/O)$ may fall because the marginal coefficient $\delta O/\delta P$ derived from equations such as those in Table 1, even if increasing over time, is more than offset by a falling P/O ratio. The ratio P/O has in fact fallen as output expanded and prices declined. This is expected with improved resource productivity—more is produced at a given price while maintaining the rate of return on resources.

Prices also influence output in the short run through the stock of productive assets, Sp, and this impact is accounted for in the Table 2 estimates.[1] The 0.155 supply elas-

---

[1] The stock of production assets was regressed in linear form on time, P_R/P_p^{-1}, and lagged production assets with the following results:

$$(6)\quad \underset{}{S_p^t} = 0.73 + \underset{(0.025)}{0.072\ (\text{Time})} + \underset{(0.009)}{0.020\ (P_R^p/P_p)^{t-1}} + \underset{(0.036)}{0.935\ (S_p^{t-1})}$$

with $R^2 = 0.990$. Standard errors are in parentheses. From this equation, the elasticity of S_p^t with respect to P_R/P_p^{t-1}, E_{sp}, was 0.228 for 1921–66, 0.598 for 1921–41, and 0.481 for 1948–66. And the adjustment coefficient for productive assets g_{sp} is $1-0.935=0.065$. In adjusting the aggregate supply elasticity for the impact of P_R/P_p on S_p, the short-run elasticity is equal to the elasticity of supply with respect to price derived from equations in Table 1 plus the elasticity of output with respect to productive assets derived from equations in ticity computed from equation (1) is the best single estimate of the short-run supply elasticity for the entire period. The results provide no clear mandate for a higher elasticity in the postwar period nor in increasing price phases; hence, 0.155 may be the best estimate for any of these conditions. Nevertheless, supply elasticities are also shown for increasing and decreasing prices and for the postwar and prewar periods. Equations (5) and (6) indicate that the short-run supply elasticity could be 0.17 for price increases and 0.07 for price decreases. Estimates from equations (2) and (3) indicate that the short-run elasticity could be 0.08 in the prewar period and 0.16 in the postwar period. But elasticities computed from equation (1) contradicted this conclusion in that a declining P/O ratio is multiplied by a fixed marginal coefficient 0.0693. Additional evidence from other techniques is examined before a final interpretation is made of the short-run supply elasticity.

One hypothesis is that the equations in Table 1 fail to provide strong support for rising elasticities over time because the decreasing prices and attendant low elasticities occurred in the postwar period. Although concentration of decreasing or increasing prices in one of the segmented time periods could cause confounding of the two major hypotheses tested—that is not the case in Table 1. Excluding 1942–47, of the nineteen years of increasing prices in the 1921–66 period, eleven occurred in the 1921–41 period and eight in the 1948–66 period. And of the twenty-one decreasing prices, ten occurred in the 1921–41 period and eleven in the 1948–66 period.

TABLE 2

ESTIMATES OF AGGREGATE SHORT-RUN SUPPLY ELASTICITY OF U. S. FARM OUTPUT FOR SPECIFIED PERIODS*

Period	Aggregate Supply Elasticity (short run)
1921–66	
Without regard to direction of Price movement (Equation 1)	0.155
Price increases (Equation 5)	.169
Price decreases (Equation 6)	.069
1921–41	
(Equation 1)	.132
(Equation 2)	.080
1948–66	
(Equation 1)	.085
(Equation 3)	.162

* Elasticities computed from equations in Table 1 at the means of the respective periods as indicated.

The long-run supply elasticity from all equations except (7) in Table 1 can be computed by considering the long-run impact of prices on Sp and T. Productivity, T, regressed on prices and a time variable revealed no association between prices and T. Thus, the impact of prices on output through T can be ignored. The influence of prices on output through Sp is found from a regression of Sp on prices and other variables as indicated previously in footnote 1. The long-run supply elasticity is estimated at 0.31 both for the 1921–41 (equation (2)) and for the 1948–66 (equation (3)) periods. Equation (1) indicates a 0.19 long-run supply elasticity for the 1921–66 period. The long-run elasticity computed from equation (7) is also in line with this estimate. These estimates are considerably below the 0.6 long-run supply elasticity estimated by Heady and Tweeten in 1963 [5, p. 435]. The long-run supply elasticities computed from the equations in Table 1 are also low when compared with results from the disag-

---

Table 1 multiplied by the elasticity of productive assets with respect to P_R/P_p, E_{sp}. The long-run aggregate supply elasticity is obtained in a similar manner except that the elasticity of productive assets with respect to P_R/P_p is divided by the adjustment coefficient for productive assets, g_{sp}, in the above calculations.

TABLE 3.

ESTIMATED AGGREGATE SUPPLY ELASTICITY OF FARM OUTPUT, WITH CROP AND LIVESTOCK COMPONENTS*

Item	Elasticities	
	Short run (two years)	Long Run (many years)
Crops†	0.17	1.56
Acreage	.04	.10
Yield per acre	.15	1.50
Livestock‡	.38	2.90
Animal units (stock)	.12	1.80
Yield per animal unit	.26	1.10
Aggregate supply elasticity§	.25	1.79

* Data interpreted from [1] [5, Chap. 16].

† The elasticity of crop production C with respect to farm price P is E_{CP}. Given the elasticity of acreage with respect to price E_{AP}, the elasticity of yield with respect to price E_{YP}, and the elasticity of yield with respect to acreage $E_{YA} = -.4$, then $E_{CP} = E_{YP} + E_{AP}(1 + E_{YA})$.

‡ Computed as in above footnote, with the elasticity of livestock yield with respect to animal units E_{YA} equal to zero.

§ Computed by the formula:

$$E_{OP} = E_{CP}\frac{C}{O} + E_{LP}\frac{L}{O} + E_{LC}\,E_{CP}\frac{L}{O}$$

where O is total farm output, C is crop output, L is livestock output, and P is the index of prices received by farmers. E_{OP} is the supply elasticity of O with respect to P. Other elasticities E_{ij} are interpreted accordingly. E_{LC} is assumed to be -0.4 in the short run and -0.6 in the long run.

gregate approach presented later. We conclude that high correlation between productive assets, Sp, productivity, T, and, in the case of equation (7), lagged output biases the long-run supply elasticity toward zero.

Aggregate Supply Elasticity Computed from Crop and Livestock Components

Estimates in Table 3 of the elasticity of crop acreage, yield, and other components of aggregate supply with respect to price are taken from previous studies. In these studies, the coefficients of livestock and crop supply functions were computed by least squares from time series. The components are weighted and combined to form an estimate of the elasticity of aggregate supply. The underlying data were not adequate for separating the elasticities for increasing and decreasing prices and for postwar and prewar periods.

For all crops, the acreage response to price is quite low in both the short and long run (Table 3). But yield response is considerable in the long run. The total elasticity of farm crop production with respect to prices received by farmers is the acreage elasticity plus the yield elasticity corrected for the negative effect of higher acreage on yields as production moves to lower yielding land.

The production period for livestock such as poultry and hogs is comparatively short; hence, animal units of livestock show a greater response than acreage to higher farm prices (Table 3). While output per animal unit is more responsive to price than is the number of animal units in the short run, the reverse is true in the long run. That is, there is greater potential for expanding livestock production by increasing the number of animal units than by increasing yield per animal unit over an extended period. Empirical results do not show that increasing the number of animal units depresses the yield per unit; hence the supply elasticity of total livestock output is the simple sum of the unit and yield components.

The total elasticity of farm output with respect to product price is a weighted sum of the livestock and crop components. There are opportunities to substitute livestock for crops by feeding. We attempt to include this interaction by a term E_{LC} which is the elasticity of livestock output with respect to crop output (see footnote §, Table 3). With a given supply of crops, an increase of livestock means less crops for final sales to consumers; hence, E_{LC} is negative. The elasticity of total supply computed from data in Table 3 is 0.25 in approximately two years and is 1.79 in many years. This implies that a sustained 10 percent increase in prices received by farmers raises farm output 2.5 percent in two years and 17.9 percent in many years. The reliability of some of the estimates such as E_{LC} in Table 3 is low, and the elasticities must accordingly be interpreted with caution.

Aggregate Elasticity of Supply Computed from Farm Input Components

Agricultural economists place much emphasis on factor markets as the roots of the supply adjustment problem in agriculture (see, for example, Hathaway [4], Heady and Tweeten [5], and Glenn Johnson [7]). And in stratifying capital inputs into groups with similar characteristics with respect to durability, degree of specialization, and source of supply, Quance found that expendables are adjusted toward rates of use which equate relevant *ex post MVP*'s or lifetime earnings with original acquisition costs more rapidly than durables, farm-produced durables more rapidly than nonfarm-produced durables, nonfarm-produced expendables more rapidly than farm-produced expendables, and inputs which are unspecialized to the farm sector more rapidly than specialized inputs [8]. These differing adjustment rates for farm inputs have important implications for aggregate supply elasticity. Before drawing conclusions concerning the magnitude of aggregate supply elasticity, a further estimate relating the elasticity of supply to the adjustment of specified input categories is obtained.

This estimate of supply elasticity is determined from the elasticities of production and the elasticities of input demand with respect to prices received by farmers. In other words, output is a function of the productivity of an input and its level of use. And the level of use is a function of farm product prices.

From the estimates of production elasticities and elasticities of demand for specified inputs in Table 4, the contribution of an input i to aggregate supply elasticity is obtained by multiplying the elasticity of production, E_{Oi}, by the elasticity of demand, E_{ip}, for input i. The

TABLE 4

ESTIMATED ELASTICITIES OF PRODUCTION, INPUT DEMAND AND PRODUCT SUPPLY FOR U. S. AGRICULTURE*

Input	Elasticity of Production	Elasticity of Demand				Contribution to Aggregate Supply Elasticity	
		Own-Price		Product Price			
		SR†	*LR*†	*SR*	*LR*	*SR*	*LR*
Fertilizer and lime	0.06	−0.6	−1.8	0.5	2.4	0.030	0.144
Machinery operating expenses	.11	−1.0	−1.5	.5	2.5	.055	.275
Feed, seed, and livestock	.09	−.8	−1.5	.7	2.0	.063	.180
Miscellaneous current operating expenses	.09	−.3	−.5	.3	2.5	.027	.225
Crop and livestock inventories	.04	−.2	−1.0	.2	2.5	.008	.100
Machinery inventory	.10	−.2	−1.0	.2	2.6	.020	.260
Labor	.25	−.1	−.5	.1	1.0	.025	.250
Real estate	.28	‡	‡	.1	.3	.028	.084
Total elasticity	1.02					.256	1.518

* Data based on [5] [9] [10] and [11].

† *SR* refers to a short run of two years; *LR* refers to a long run of many years.

‡ Nearly zero.

total price elasticity of supply of farm output E_{CP} is then

$$E_{CP} = \sum_{i=1}^{n} E_{Oi}E_{ip}$$

where there are $n=8$ input categories.

A 1 percent increase in use of fertilizer increases farm output 0.06 percent, whereas a 1 percent increase in farm real estate increases farm output 0.28 percent. Because farm real estate tends to be quite fixed, its large impact on output is offset by its low response to farm product prices; hence, fertilizer contributes more to the aggregate supply elasticity than does farm real estate.

The total elasticity of supply is 0.26 in two years and 1.52 in many years, Table 4. But these results apply primarily when farm prices are increasing. Fixity of assets tends to reduce the supply elasticity applicable to falling farm prices. In periods of decreasing product prices, farm output is contracted mainly through reduction in operating inputs directly, and indirectly through reducing the flow of services from economically fixed durables complementary with the reduced operating inputs. If only operating inputs in Table 4 are variable when prices fall, the short-run elasticity is reduced from 0.26 to 0.18, and the long-run elasticity is reduced from 1.52 to 0.82. If product prices fall sufficiently low and for a sustained period, the stock of some fixed durables would be depleted, not replaced, or salvaged, and the supply elasticity would rise accordingly.

Summary

The three approaches—(1) direct least squares, (2) separate yield and production unit components for crops and livestock, and (3) separate input contributions—used to compute the elasticity of farm output with respect to prices received by farmers indicate a low but decidedly positive supply response.

Point estimates of the aggregate supply elasticity computed in this study fall in the 0–2 range—certainly not a wide range in total perspective. But the disparity in estimates does require considerable interpretation. And it is difficult to give a precise answer to the policy-maker who wishes to know the response of farm output to price.

In interpreting all the above evidence and weighing it according to our judgment of the reliability of data and methodology, we provide the following conclusions assuming that we are pressed to give point estimates of the aggregate supply elasticity. No basis was found for distinguishing between the prewar and postwar elasticity estimates but the following elasticities are specifically intended to reflect the postwar period.

The supply elasticity is 0.10 in the short run and 0.80 in the long run for decreasing prices. But the supply elasticity is 0.15 in the short run and 1.5 in the long run for increasing prices.

Based on these estimates, a once-for-all decrease of 10 percent in prices received by farmers would decrease output 1 percent in two years and 8 percent after many years. But a once-for-all increase of 10 percent in prices received by farmers would increase farm output 1.5 percent in two years and 15 percent after many years. The relative difference is less between the short-run elasticities for the two price phases than between the long-run elasticities. This is because the operating inputs which generate the short-run response are consumed with little carryover of service; hence, give rise to little resource fixity and to greater symmetry in response to price. The long-run elasticity, which depends on durable inputs, is more heavily influenced by resource fixity and this is reflected in the considerable differences in long-run elasticities for increasing and decreasing prices.

References

1. H. E. Buchholz, G. G. Judge, and V. I. West, "A Summary of Selected Behavioral Relationships for Agricultural Products in the United States," Dept. of Agric. Econ., AERR-57, Univ. of Illinois, 1962.
2. Willard Cochrane, "Conceptualizing the Supply Relation in Agriculture," *J. of Farm Econ.*, 1955, pp. 1161–76.
3. "Economic Report of the President" (U.S. G.P.O., Jan., 1967).
4. Dale E. Hathaway, *Government and Agriculture* (Macmillan Co., 1963).
5. Earl Heady and Luther Tweeten, *Resource Demand and Structure of the Agricultural Industry* (Iowa State Univ. Press, 1963).
6. Earl Heady, *et al.*, eds., *Agricultural Supply Functions* (Iowa State Univ. Press, 1961).
7. Glenn L. Johnson, "Supply Functions—Some Facts and Notions," *Agricultural Adjustment Problems in a Growing Economy* (Iowa State Univ. Press, 1958).
8. C. Leroy Quance, "Farm Capital: Use, M.V.Ps and Capital Gains or Losses, United States, 1917–1964" (unpublished Ph.D. thesis, Michigan State Univ., 1967).
9. Luther G. Tweeten and Ted R. Nelson, "Sources and Repercussions of Changing U. S. Farm Real Estate Values," Okla. Agr. Exp. Sta. Tech. Bul. T-120 (Stillwater, 1966).
10. Fred H. Tyner and Luther G. Tweeten, "A Methodology for Estimating Production Parameters," *J. of Farm Econ.* 1965, pp. 1462–67.
11. ———, "Optimum Resource Allocation in U. S. Agriculture," *J. of Farm Econ.*, 1966, pp. 613–31.
12. U. S. Bureau of the Census, *Historical Statistics of the United States* (Washington, D. C., 1960).
13. U. S. Dept. of Agric., "The Balance Sheet of Agriculture," Agr. Inform. Bul. No. 329 (Washington, D. C., 1967), and preceding issues.
14. U. S. Dept. of Agric., "Changes in Farm Production and Efficiency," Statis. Bul. No. 233 (Washington, D. C.), June, 1967, and preceding issues.

DISCUSSION

MARVIN KOTTKE: Sharples took us on a tour of the RFA model, and I want to commend him on the way he conducted the tour. However, after giving us a critical view of the RFA model, he suggested we take the long-run road on our next excursion. As one interested in taking the next excursion, I take issue with his suggestion because I see the long-run supply model as a dead-end road.

In laying the groundwork for my argument, I shall first talk about linkages. In the past decade, as pointed out by Sharples, considerable effort was devoted to finding a linkage between micro and macro supply response by use of the RFA model. The Tweeten-Quance model does not seek a micro-macro linkage but it starts at the opposite pole and attempts to disaggregate the macro in the direction toward micro units. I would submit, however, that the micro-macro linkage is only one of several that need closer attention if we are to make further advances in supply research. Time and spatial linkages are two more that are extremely relevant in contemporary supply behavior. They are especially so because of the current interest in building models for realistic application. In reality, the whole supply problem is intricately interwoven with time, spatial and micro-macro linkages. Our job is to find the best way to slice the problem into manageable relations and still capture the linkage effect.

It is clear that we have been less than successful in finding reasonable micro-macro linkages. But it may be of some comfort to know that we are not alone in the search. A year ago at our Association's meeting,[1] a session was held on micro-macro linkages in planning a rural economy. The linkage seemed to be elusive in that session also. Even its usefulness was doubted. In contrast, the linkage in supply response would not only be useful, but necessary. In supply analysis, the micro-macro linkage is primarily a feedback effect. We need to capture not only the individual firm's response, but also the feedback to the firms of the effect of their combined responses. It seems to me that a model incorporating cobweb and recursive relations could accomplish this type of micro-macro linkage.

The search for a spatial linkage has been more successful. Spatial linkages can now be handled with quadratic programming models for perfectly competitive industries. We still need to explore ways of introducing imperfectly competitive conditions. In the Northeast Dairy Adjustment project, we have made a preliminary test of a spatial model incorporating imperfect competition and institutional conditions and the results are promising. In using quadratic programming we have learned that the reasonableness of intercept parameters—their magnitude and their change through time—are just as crucial as elasticities. Hall, *et al.*,[2] made the same observation in their application of quadratic programming. As a consequence, I expect that, in the future, we will be putting more effort into testing the reasonableness of the intercept parameters than in estimating elasticities.

Our attempts at handling time linkages in supply response have been rather vague in the past. Usually we have specified a length-of-run for the time dimension. But lengths-of-run, especially intermediate and long run, are incapable of handling time linkages for the following reasons.

First, in reality, there is no single length-of-run that would represent all firms' supply behavior at a given point or period in calendar time. One of the shortcomings of the RFA model was that all farms were assumed to respond to price in unison during a particular calendar period of time. If one abstracts from calendar time, it is conceptually logical to assume that all firms could respond in unison and make the same length-of-run decision.

---

[1] K. D. Cocks and H. O. Carter, "Micro Goal Functions and Economic Planning," *Amer. J. of Agric. Econ.*, May, 1968, pp. 400–411.

[2] H. H. Hall, E. O. Heady and Plessner, "Quadratic Programming Solution of Competitive Equilibrium for U. S. Agriculture," *Amer. J. of Agric. Econ.*, Aug., 1968, pp. 536–55.

But, if one dates supply response with calendar time, it is not logical to assume that all firms respond in unison and in the same direction. At any given calendar point in time some farmers may respond positively, others respond inversely, and still others do not respond at all. The differences in each firm's supply behavior are related to structural differences among farms, including such things as rate of technological adoption, capital position, and stage of life.

Second, the theoretical basis of the long-run supply model, particularly in the normative approach, has its roots in the neoclassical cost curve analysis. Viner's abstract, long-run supply curves which explain increasing, decreasing, or constant cost industries, are quite inept for dealing with supply response in the 1960's. A clue to the long-run supply model's inappropriateness for today's agriculture can be detected in Viner's[8] definition of long run, "a sort of 'timeless' long run throughout which nothing new happens except the full mutual adjustment to each other of the primary factors existing at the beginning of the long-run period." This definition clearly rules out technological, institutional, and price changes through time. By definition, then, it cannot deal with the major variables in agricultural adjustment; namely, technological change and capital investment.

Third, use of a long-run supply model introduces a semantics problem. Long-run supply has different meanings in different contexts. It means a relatively rigorous static supply curve in the neoclassical sense. It means a "mongrel" type of supply estimate in the loosely defined sense. It means a time-oriented supply relation in the dynamic sense. To distinguish two of these different meanings, supply researchers use the terms "normative" long-run supply for the RFA type of model and "positive" long-run supply for the time series regression type of model. Unfortunately, these two terms have probably generated even more confusion. Nerlove's dynamic long-run supply response differs in meaning from the others in still another way. The incorporation of distributed lags results in a mixture of dynamics and long run—a combination that produces mixed meanings. The proliferation of meanings seems to stem from a desire to stretch the long-run supply model's use beyond its original capacity.

The time has arrived to retire the long-run supply model. In its place I suggest we use comparative statics and dynamic supply models. By use of a comparative statics framework, relations involving technological change, capital investment, institutional change, and stage of life can be brought to bear on the supply estimates through the shift parameters. The comparative static points can be tagged with calendar time without much difficulty and the linkages in time can be accomplished with the dynamic concepts of the cobweb effect, recursive relations, and lagged variables.

In conclusion, I agree with Sharples that we need redirection in supply research, but I say we should take a road which offers micro-macro, spatial and time linkages in a semi-dynamic framework.

W. Neill Schaller: Tweeten and Quance demonstrate imagination in their use of the aggregate production function to measure national supply response for all commodities. While not denying the importance of supply elasticity at such a high level of aggregation, I suspect the authors agree that its usefulness would be enhanced if it could be combined into a larger package of information. I wonder if it would not be possible, now, to develop a consistent, integrated system of commodity supply elasticities at the U.S. level, including cross-elasticities—like the demand system developed by Brandow.[1]

I should think that the resulting package would contribute more to policy formulation than separate unintegrated studies, and that the system would provide a means of checking for internal consistency.

Elasticities have a long-standing conceptual and numerical meaning to the economist. But it is always difficult to prevent their misinterpretation. Tweeten and Quance recognize the problem. Still, I am not sure that they overcome the likely confusion between economic and calendar time when they define short run as "two years" and long run as "many years."

---

[8] J. Viner, "Cost Curves and Supply Curves," *Zeitschrift für Nationalokonomie* III (Verlag von Julius Springer in Wien and Berlin), 1932, p. 29.

[1] G. E. Brandow, *Interrelations Among Demands for Farm Products and Implications for Control of Market Supply*, Penna. Agr. Exp. Sta. Bul. 680, Aug., 1961.

The Tweeten-Quance paper provides several relevant results, including a courageous offering of point elasticities. They tell us that "there is no basis to conclude that the structure of supply changed between the postwar and prewar periods." They see little basis to conclude that output response to price has increased over time, or that the response differs for rising and falling prices. I had trouble reconciling these remarks with the different elasticities for increasing and decreasing prices, presented at the end of the paper. I suspect that my "hang up" is explained by the word "structure." The authors' results support the hypothesis that the structure of supply was different for increasing and decreasing price years. But what is this "thing" called structure? Is it captured at such a high level of aggregation?

The approach discussed by Sharples is quite different from the one found in the Tweeten-Quance paper. Sharples demonstrates that it is indeed possible to disaggregate total supply, geographically and commodity-wise, while maintaining an aggregative point of view. Not many years ago, supply analysts seemed to be clearly either micro-oriented or macro-oriented. The distinction has grown fuzzy. This I regard as a healthy sign.

Yet, we are still a long way from capitalizing on the complementarity of research techniques and of analyses cast at different levels of aggregation. I agree with Sharples that the micro-macro linkage may have to be informal. In fact, our search for a formal linkage—though scientifically appealing—may explain some lack of progress on achieving the desired complementarity.

It is perhaps significant that neither paper includes a detour into the semantics of normative and predictive research. This, too, is an encouraging sign.

The results presented by Tweeten and Quance obviously apply to output and price levels that have been experienced. Similarly, the estimates discussed by Sharples are determined in part by the range of historical experience. We all know what we are up against when we go beyond that range. But as others have surely said before me, some of our most pressing policy questions require estimates of supply response to conditions not yet experienced. Consider, for example, questions such as: (1) How would output respond to substantially lower prices? (2) How much new cropland would come into production if a general cropland retirement program were enacted? (3) What would be the effects on production and income if payment limitations were imposed? These questions are not academic. Unfortunately, our tools seem better suited to questions that involve relatively insignificant departures from the current response environment.

The papers presented here tell me that we have made good progress since 1960 when Nerlove and Bachman wrote their important survey article on supply estimation.[2] Useful modifications of existing tools have been developed, and our research has become more mature. But is it possible that the marginal returns to improved use of traditional techniques are now well below the returns that could accrue to research on approaches that lay idle, or do not exist, in our tool kit?

For example, Nerlove and Bachman talked about the need to undertake and extend the use of the producer panel device. What has been done along this line? Could producer panels help answer questions of the kind posed above—those that take us outside of historical experience?

Several years ago, when a relative of mine asked me what my research involved, I explained that I was using mathematical techniques to estimate how farmers would respond to changes in government programs, prices, and so on. Unimpressed, my relative replied, "If you want to know what farmers would do, why don't you ask them?" The question has nagged me ever since. To my knowledge, we have not really studied the relationship of producer plans to their expectations, nor the relation between their plans and actual behavior. Is the theoretical foundation lacking, or are we somehow tied—intellectually or organizationally—to the kinds of approaches discussed in this session?

---

[2] Marc Nerlove and Kenneth L. Bachman, "The Analysis of Changes in Agricultural Supply: Problems and Approaches," *J. of Farm Econ.*, Aug., 1960, pp. 531–54.

JAMES H. WHITE: I will comment briefly on the Quance-Tweeten and Sharples papers and then consider some concepts of agricultural supply and problems involved in estimating supply functions for agricultural products. Some of my remarks will be directed toward the degree to which we have been conscious of these problems and to possible reasons for not taking them into consideration in the analytical models employed in studies relating to agricultural supply. We should be cautious about making final judgments on the ultimate value of our research efforts in estimating agricultural output. Agriculture is a dynamic industry and to determine exactly what it will look like at any point in the future is not possible.

The stated objective of the Quance-Tweeten paper was to develop positivistic measures of aggregate agricultural output elasticities. They visualized that knowledge of the aggregate output response to changes in economic conditions would be helpful to those responsible for "public policies concerned with the earnings of all farm products and total farm income. . . ." Estimates of the aggregate supply function and associated elasticities were made using three approaches: (1) direct least squares, (2) separate yields and production unit components for crops and livestock, and (3) separate input contributions to total agricultural output. There was a tone of disappointment by the authors in the results of their analysis. They stated that it is difficult to give a precise answer to a policy-maker who wishes to know the response of output to price but made concluding statements based on their judgment of the reliability of the data and the methodology. This is not necessarily bad. The position most of us find ourselves after any research effort is something short of complete knowledge of all the relevant variables and how they relate to each other and to the problem under consideration. A student of mine once remarked that if we publish only that which is statistically significant at the 1 or 5 percent level of probability, we would have a small printing bill.

Although estimates of aggregate agricultural output elasticities may be useful for some purposes they are "too aggregate" for making inferences to lower levels of aggregation. It is not possible to separate out the impact of price changes by regions, by type of farm, by commodity, or by the magnitude of the various inputs used in production. However, the results obtained by Quance and Tweeten possibly could be made more useful by including in the paper more interpretation of the results as they indicated the need for doing.

Sharples, as well as others,[1] have done an excellent job in discussing the role representative farms have played in estimating aggregate supply response. Nothing is to be gained by heaping criticism on the approaches that have been taken in supply studies. Virtually all of the problems encountered have long been recognized and were reemphasized by Plaxico and others[2] before most of the recent supply studies were initiated. Possibly too much was expected of the models we were using and too little effort was made in developing meaningful coefficients.

The prime reasons we committed ourselves to the representative firm approach for studying aggregate supply and/or output response in agriculture were because Alfred Marshall legitimized the approach and it has been paramount in the training of most agricultural economists. I am of the opinion that the inclination for economists to continue using this approach has its genesis in their formal training where emphasis was placed on profit motivated static economic theory of the firm and its linkage to the industry. Then, too, the tools at our disposal—linear programming, spatial equilibrium analysis, regression, and others—were thought to be potent tools for quantifying input-output relationships of the firm. Had we taken a course that would have led to relating the motives of managers, their age distribution, their equity position and other less quantifiable subjective aspects of the firm and firm-household relationships of the farm to output response, we would have been banished from the profession. As viewed by some the results from such an approach probably would have been no more useful than the results obtained by the representative farm approach. In view of the above it should come as no surprise to anyone that agricultural economists took the representative farm approach to the study of aggregate supply.

---

[1] Harold O. Carter, "Representative Farms—Guides for Decision Making?" *J. of Farm Econ.*, Dec., 1963, pp. 1448–55.

[2] J. S. Palxico, "Aggregation of Supply Concepts and Firm Supply Functions," *Farm Size and Output Research*, So. Coop. Ser. Bul. 56, pp. 76–91, and 112–114, June, 1958.

The many facets of supply may be the source of some of the difficulty and confusion economists have encountered in estimating supply and using supply estimates foreign to the purpose for which they were developed. Evidenced by the two papers presented here, estimates of agricultural supply can relate to estimating the response of total agricultural output of a nation, or to the response in output of a single product for a given producing unit to changes in product prices. Supply studies can also relate to the classical supply relation that refers to how the quantity offered for sale varies as its price varies relative to other product prices for some given time period and level of technology. Estimates of supply may refer to what Cochrane[3] calls an output response relation describing what will happen to the commodity offered for sale when the time period and technology are not held constant. Supply estimates can relate to: static or dynamic situations; short-, intermediate- or long-run time periods; commodity or aggregate agricultural output; national policy formulation or guidance to farmers; single or multiproduct firms; a farm or a region; and others. A step toward avoiding some of the difficulties we have encountered in supply studies is to explicitly specify the purpose for which the supply estimates are made, the assumptions, the source and reliability of the data, and methodology compatible with the data and the objectives of the study.

Effort is likely to continue in estimating the short- and intermediate-output response of firms to changes in physical, economic, and institutional conditons. The exact methodology that will prevail is not clear. There are some who are yet to be convinced that the potential of the representative farm approach has been fully exploited. Regardless of the approach the ultimate test will be the usefulness of the results. In all probability the reward from these efforts will be in proportion to the effort made in problem identification and specification, the development and refinement of technical coefficients and other relevant data, and the choice of model(s). Attention must be directed toward these things if we hope to duplicate reality—become more predictive—with estimates of how the output from agricultural firms changes in response to changing conditions.

---

[3] Willard W. Cochrane, "Conceptualizing the Supply Relation in Agriculture," *J. of Farm Econ.*, Dec., 1955, pp. 1161–76.

PROBLEMS IN THE THEORY OF PUBLIC CHOICE

SOCIAL COST AND GOVERNMENT ACTION

By Gordon Tullock
Virginia Polytechnic Institute

In Illinois where I was born—and indeed in most of the two American continents—the common mosquito is a major pest. With the development of DDT during World War II, a cheap way of reducing the mosquito population became available, and a very large number of communities in the years after World War II hired aircraft to fly low over the city and spray it with DDT. Recently this has become much less common. The reduction in the amount of such spraying is partially the result of a realization that the DDT spray has other effects on the natural environment than the reduction of the mosquitoes, and that some of these other effects may be quite undesirable, and partially the result of home air conditioning. A person living in an air-conditioned home is unlikely to spend as much time outside as a person living in a non-air-conditioned house. Although the amount of aerial spraying of mosquitoes has been considerably reduced in recent years, a number of other techniques which are both less effective in dealing with mosquitoes and more expensive have been adopted in some areas. They are not as widespread as aerial spraying was, let us say, ten or fifteen years ago.

The reader may wonder why a discussion of social cost should begin with this technological discussion of a rather unusual local problem. The reason is simple. Almost all of the problems involved in decisions as to what types of activity should be undertaken by the government will be found in this simple example. Further, there is absolutely nothing in the way of a traditional solution to this problem. One of the greatest problems in talking about the new discoveries in the field of social cost, externalities, and what we might call the economics of the government sector is that most people have learned, normally very early in life, the existing tradition. Those who have not accepted the existing tradition normally learn some particular attack on it which is in many ways just another tradition. Thus discussion of government activities runs instantaneously into a barrier of very strongly held ideas. If we discuss mosquito abatement, however, we normally find a complete absence of these traditions or antitraditions and hence can deal with the problem with less emotional difficulty.

To return, then, to the mosquitoes, let us assume that we are in 1952 and that we are living in a small town in Iowa where our mosquito problem is bad. We could undertake a spraying campaign in our own backyard. We could buy in a local supermarket a suitable insect spray and spray our backyard and perhaps spray a little into the neighbor's yard. This spray would in general reduce the number of mosquitoes in the area, partly by killing the ones who were there at the time we sprayed and partly because the spray left some residium on the vegetation and mosquitoes landing on the residium might be killed. Nevertheless, this would be a very expensive and not tremendously efficient way of reducing the number of mosquitoes.

Another method of mosquito abatement was available, however; for $50.00 we could hire an airplane and this would spray the entire town with a density to give me or any other citizen considerable mosquito abatement. This method, like

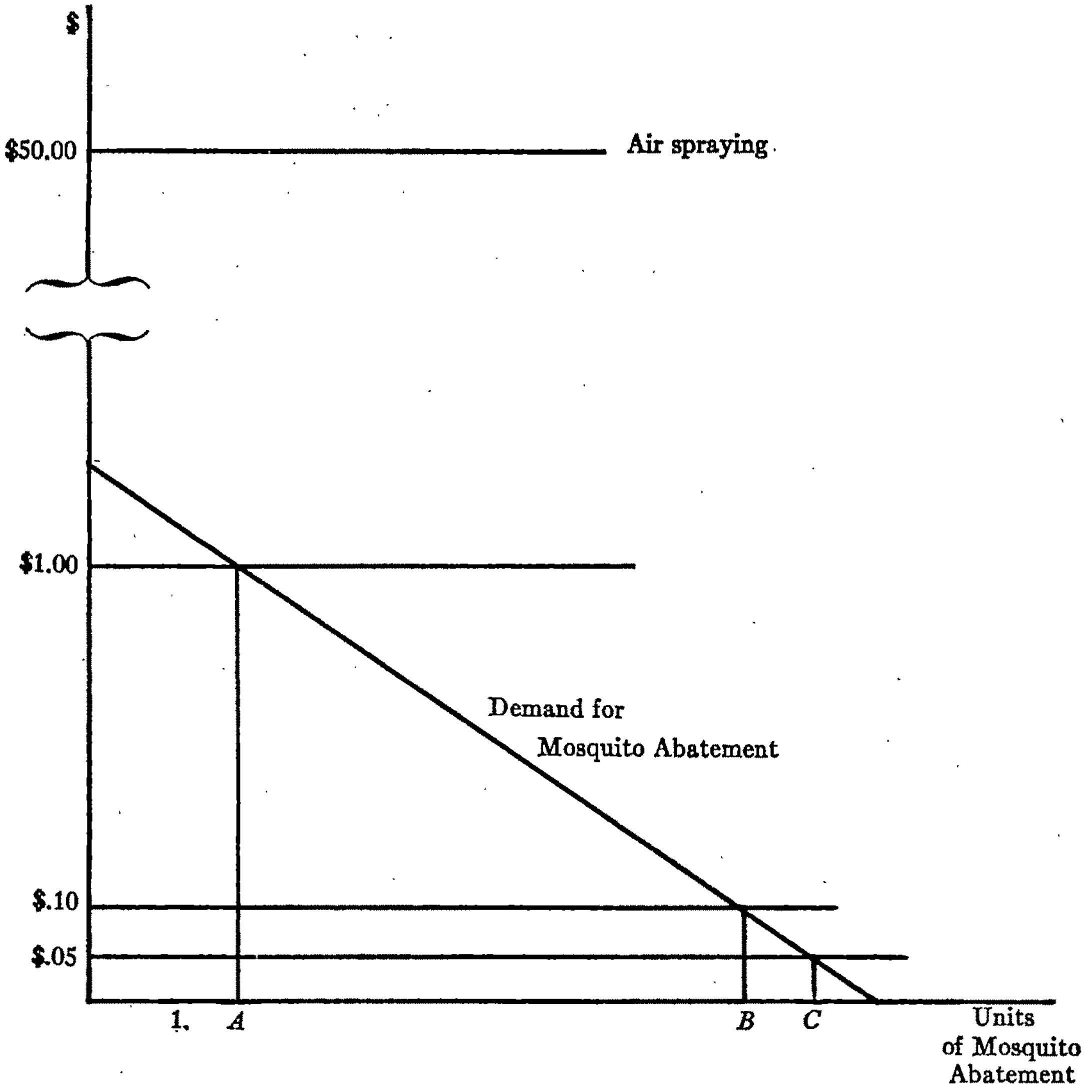


FIGURE 1

the use of handsprays, could be intensified with additional deaths of mosquitoes. The situation is shown on Figure 1. For some individual the demand for mosquito abatement is shown by the usual slanting line. If he chooses to use the handspray method, then the line marked $1.00 will indicate the cost of killing "one unit" of mosquitoes. The individual will choose to purchase *A* units, and the total cost he will run for this will be represented by the rectangle to the left of *A*. The individual will not be interested in hiring an airplane which, from the standpoint of the individual, is completely dominated by handspraying. He can obtain any amount of mosquito abatement in his own yard that he desires more cheaply by handspraying than from the air.

The situation changes radically, however, if the individual clubs together with the other citizens of the town to hire the airplane. Suppose there are 1,000 citizens in the town and they get together to hire the plane. The cost to each of them for mosquito abatement in his backyard falls to five cents a unit and the total amount consumed would rise to *C*. Under these circumstances the entire mosquito abatement demand would be taken up by air spraying rather than handspraying because, once again, there is strict dominance of this particular technological method.

Clearly, the individual if given a choice would prefer *C* mosquito abatement by air spraying to consuming *A* units by handspray. Thus there would be a good argument for collective provision of mosquito abatement. Suppose, however, that instead of having some government instrumentality undertake mosquito abatement, an effort is made to have various individuals privately contribute money to hire an airplane. Under these circumstances the cost of hiring an airplane depends on the number of people who have contributed. If only 500 people are willing to contribute, then the cost of mosquito abatement by air (per unit) would be ten cents and they will choose to purchase the amount *B*.

Assuming that this voluntary method of purchasing mosquito abatement is adopted, the individual would be rational not to make his ten cent payment. If he is not a member of the group making the payment, he receives the mosquito abatement free. If, on the other hand, he decides to make his payment and we assume that the amount of money he puts in is then invested in purchasing additional aircraft time for the whole city, then he faces a price for purchasing mosquito abatement in his yard of $50.00 a unit. This price is clearly way above the amount that he wishes to pay. It is extremely unlikely that individuals would be willing to make voluntary contributions. Normally only a government could provide the airplane spraying.[1]

We thus have what appears to be a fairly unambiguous argument for a governmental agency compelling the citizens of this small town to make the five cent payments for the hire of an airplane. The citizens themselves would be better off under this arrangement and would presumably favor it. There has, however, been an implicit assumption in the discussion so far—which is that each of the citizens has exactly the same demand for mosquito abatement. Presumably, this is not true. This means that a decision must be made as to how much mosquito abatement should be purchased. In order to consider this decision, let us shift to Figure 2 in which we show the demand curves for mosquito abatement of three citizens (Mr. A, Mr. B, and Mr. C). Note that if there is no decision to hire an airplane, the three individuals will simply purchase different numbers of units of mosquito abatement through the use of the handspray: Mr. A purchasing a' units, Mr. B puchasing b' units, and Mr. C purchasing c' units.

If, however, it is decided to hire an airplane and engage in collective provision of mosquito abatement, then some kind of decision must be made as to how much mosquito abatement should be purchased. It will be noted that the three individuals have different ideas as to how much this should be—represented by *a, b,* and *c* on the diagram. I have carefully constructed this example so that their preferences on this particular point would be what is known as "single peaked."[2] If we consider

---

[1] Sometimes informal pressures can function very much like a government. As a general rule, however, human experience seems to indicate that informal pressure is not sufficient and we normally use governmental pressure in such cases.

[2] Cf. Duncan Black, *The Theory of Committees and Elections* (Cambridge, 1958).

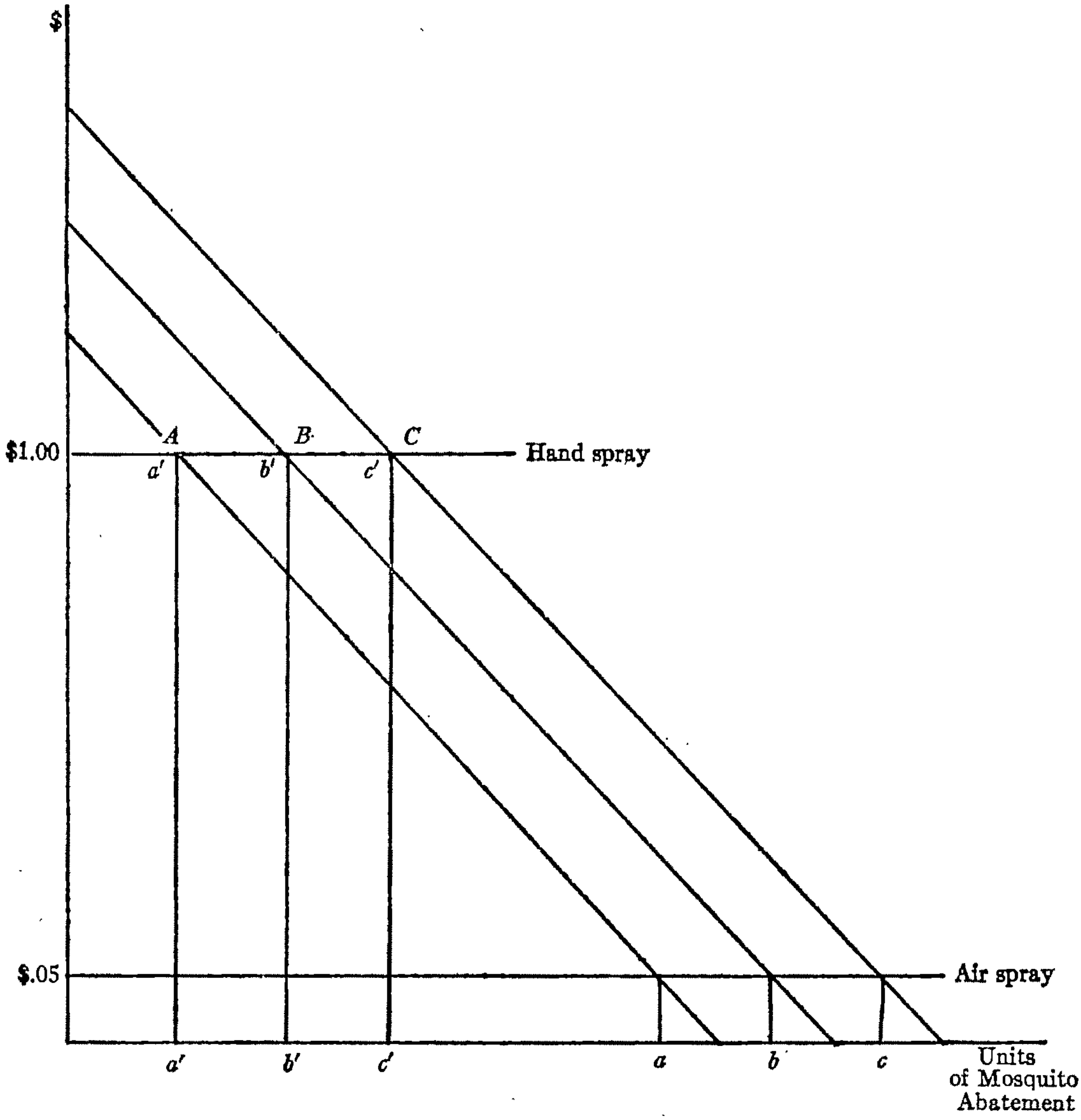


FIGURE 2

ourselves as dealing with only a three-man community and they make decisions by majority voting, and in this particular type of situation there are arguments for doing this, then we would predict that they would purchase the amount *b* of mosquito abatement. This means that Mr. A and Mr. C have failed to obtain their optimum amount of mosquito abatement.

It does not, of course, follow from the fact that the individuals could make a perfect adjustment of how much mosquito abatement they wish to purchase if we use handsprays and must anticipate in most cases that they will not obtain their exact optimum amount of mosquito abatement if we use airplanes to spray, that handspraying is superior. It merely follows that there is a cost involved in the air spraying which should be taken into account. Assuming again that we have our society of 1,000, I would compare my desire for handsprays (that is, how much I would purchase in the way of mosquito abatement by hand) with what I thought was the likely outcome of the voting process in terms of the amount which would be obtained by collective provision. I would anticipate that the collective provision would not turn out to be exactly the amount that I wanted at that price. It would not follow from this that I would favor exact adjustment under what we might call market provision to imperfect adjustment under collective arrangements. Indeed, I presume most people in areas where mosquitoes are bad if confronted with the particular problem that I have outlined would choose the collective provision. Note, however, that this means that they are choosing a less than optimal arrangement of the resources by their own preference ordering. There is, in a sense, an externality imposed upon them by the choice of the collective decision process. They will no longer be able to make an ideal adjustment.

Let us now inquire what is the ideal size of the government unit which deals with mosquito abatement. First, we should note the limits that are placed upon this by the technology of aircraft spraying. In order to be efficient, the aircraft should spray the entire town and certain nearby mosquito breeding areas. An effort to spray half of the town would give much less than half the protection. In other words, it would be a very bad bargain. Therefore, the minimum size of government units which decides to hire the airplane would be our small town in the Middle West.

By rather similar methods we can determine the maximum size of government units for mosquito abatement. In general, every increase in the size of the unit reduces the likelihood that the final provision will be very close to the desires of a given citizen. This is particularly true in something like mosquito abatement where different communities presumably have different levels of mosquito infestation. In general, as the size of government units is increased the number of externalities internalized is increased but the adjustment of the government activity to the desires of any individual voter is decreased. In our particular case, mosquito abatement, the laying off of these two factors leads pretty unambiguously to the view that the small Iowa community should provide its own mosquito abatement. With other problems, of course, other solutions should be expected.

So far, we have been talking about the problem of mosquito abatement about ten years ago. There has been considerable technological change since that time. Let us confine ourselves to considering only those discoveries which indicate that simple airplane spraying of DDT is not a desirable way of dealing with the problem. It has been realized that there are a large number of secondary costs from this operation and that these secondary costs may well be much in excess of the benefit to the spraying process. As a result, the technology of mosquito abatement no longer mainly depends on this very cheap method of killing mosquitoes. We need not go into the more complicated and more expensive methods that are now in general use. It is perfectly possible that tomorrow someone will invent another method of getting rid of mosquitoes that is as cheap as aerial spraying of DDT was thought to be when it was widely used, but we can simply note that present methods are expensive and inquire what effect this would have on the reasoning so far.

The first possible effect of the increase in expense of mosquito abatement by collective measures might be that the unit cost of a given amount of mosquito abatement by collective measures would be equal or higher than if one restricts mosquito abatement to the private use of sprays in one's own backyard. In this case, which is an easy one, the proper decision of course would be to abandon completely all collective efforts to reduce mosquitoes. The second possibility (also very easy) is that mosquito abatement either by public or by private means might become so expensive that it would be no longer desired by individuals. Here, again, the proper solution is to have no public program for mosquito abatement, and we would also anticipate that there would be no private abatement either. Both of these are easy problems and in both cases we need go no further with our analysis.

The interesting question, however, is what would we do if the use of various public means for reducing the mosquito population (let us say, specialized treatment of breeding areas of mosquitoes) is still a less expensive method of obtaining a certain amount of mosquito abatement than is private spraying, but that the difference becomes small. On Figure 3 I have drawn in this problem.

We assume here that methods of mosquito abatement by collective means exist and these are efficient enough so that if all members of the community are compelled to contribute the cost of purchasing one unit of mosquito abatement it will be \$.95 per head, whereas private purchase of one unit of mosquito abatement would remain at \$1.00. If we consider only Mr. B, clearly collective provisions would be desirable. He would be better off purchasing x units of mosquito abatement at \$.95 instead of purchasing b units at \$1.00 which is his market economy alternative. His net benefit is measured by the areas shaded horizontally and slanting to the left in Figure 3.

If, however, we consider a community consisting of three members (Mr. A, Mr. B, and Mr. C), the situation is more complicated. Mr. A, for example, benefits from the establishment of the new level of mosquito abatement to the extent of the horizontally shaded trapezoid in the upper left. He is injured through the necessity of buying more mosquito abatement than he wants even at the new price to the extent of the vertically shaded triangle. Clearly, he is much worse off with collective provision than he would be with individual provision. Mr. C is affected in a somewhat more ambiguous way. His gross gain is the gross gain of Mr. B plus the little dotted triangle. He, however, suffers a loss to the extent of the shaded triangle from not being able to purchase the additional mosquito abatement privately. This loss will only be suffered if it is not possible for him (for technological or for legal reasons) to supplement the public provision at the same cost as he could have bought mosquito abatement privately before. If the public provision of mosquito abatement actually reduced the cost of additional mosquito abatement (which is conceivable), he might gain. If the public provision did not completely bar private but simply made private supplement rather inefficient (which I imagine is the common case), then Mr. C would face a supply curve somewhat like the horizontal dashed line and would purchase $c' - x$ amount of private mosquito abatement. The gross cost to him of the new arrangement then would be the trapezoid lying between this horizontal dashed line and the \$1.00 line. If this were a smaller area than the rectangle to the left of the new supply quantity, he would gain in net terms.

Now the question arises naturally whether it would be desirable to undertake public provision of the mosquito abatement granted that we have this small three-person community. If the provision of public mosquito abatement injures both A and C and benefits B, then the ordinary

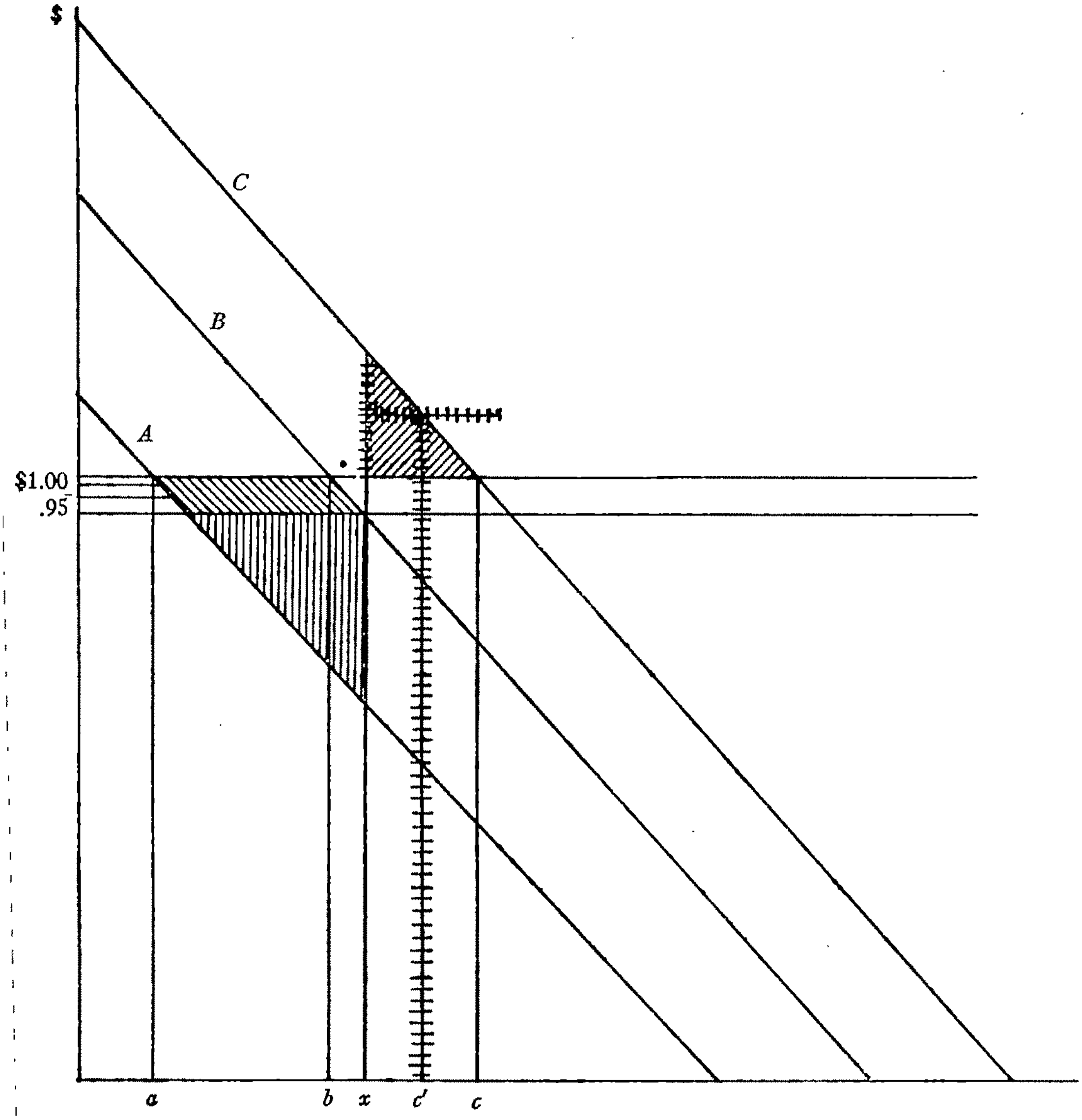


FIGURE 3

welfare economist can inquire whether B is able to compensate A and C for this injury. If (and this is also quite possible) the provision benefits both B and C but injures A, then, of course, it will be more likely that compensation can be undertaken. Unfortunately, when we are talking about public goods and where the quantity of public good consumed is set by voting or any other collective process, compensation becomes almost impossible in the real world.

The problem is that we will be unable to find out the amount of the compensation. Mr. A probably has not given much thought to how much he would require to compensate him for a combination of *x* provision collectively and the appropriate taxes. If asked to give thought to this rather strange problem, he has absolutely no motive to correctly interpret his own feelings. This would be particularly true if we were dealing not with our small three-man community. A might be concerned with the prospect that too large a claim for damages on his part might make the whole project impossible. Thus, there will be some limit to the degree to which he would exaggerate the loss which he would suffer. Similarly, B and C even if they are benefited probably would be hard pressed to put a monetary value on their benefit if we asked them and have substantially no motive to do so accurately.

As a general principle, attempting to get voters voting on some issue to compensate each other out of their individual surpluses for their individual losses is not a feasible political proposition. Thus, we cannot use the traditional welfare economics method of making a direct payment from the people who gain to the people who were injured. There is, however, another and rather debatable tool in the welfare economist's tool kit. Some (but not by all means all) welfare economists would argue that if we could compute that the payment would be possible, then it is not necessary to make it. Under this line of reasoning if there is a net social benefit, we don't need to concern ourselves with the way in which it is distributed. This method of applying the Paretian criterion is controversial and I do not wish to endorse it here. There is, however, a variant on it which is clearly a respectable rule and one which we will use.

According to this variant, if we anticipate making a large number of decisions in the future and if we cannot tell who will be benefited and who will be injured by each of these collective decisions but anticipate that most members of the society will find themselves benefited sometimes and injured sometimes, then the rule of simply computing whether there is a net benefit or not and using that for all of these decisions would probably give to each individual in society a positive discounted future income stream. It could be said under this argument that if we found a "net gain" we could undertake the collective provision of mosquito abatement without worrying about the fact that some people (particularly Mr. A) are injured. It will be noted that the use of this argument involves an income transfer. Mr. A is injured, Mr. B is benefited, and depending on the particular parameters of the problem Mr. C may be either benefited or injured. This transfer is clearly not something we would positively favor. That is particularly true since as a general rule most members of society who are interested in restricting the consumption of any particular good are apt to be poorer members. Thus transfers of the sort we are now talking about are apt to be transfers from the poorer to the better-off members of the community. Poor Mr. A is made worse off, middle-class Mr. B is benefited and upper-class Mr. C may gain.

It will be noted that the basic parameters which have led us to choose either collective or private provision of mosquito abatement have been essentially technological. It is, therefore, sensible to pause briefly and inquire exactly what the nature of technological superiority of collective provision is. At first glance, one might think that it was simply an example of an ordinary scale economy, but this is clearly not so. General Motors has surely exhausted all the scale economies that are available in the manufacture of Chevrolets, yet we find no need for collective provision here. General Motors can sell its cars to people scattered all over the United States without worrying very much about whether the next-door neighbor of any given purchaser of Chevrolet has a Ford.

The special characteristic of aerial spraying of mosquitoes is that it is generally impossible to do it economically on one city lot at a time. In order to get any economy into the operation at all it is necessary to spray a fairly broad area. This is partly because the plane must fly from its airport to the place where it releases the spray anyway and partly that the characteristics of the spray are such that it is apt to cover several surrounding house lots anyway. If only one of them is paying for it, the others will receive a free ride. The problem, then, is geographical contiguity. Geographical contiguity is a basic characteristic of almost all such areas where we would choose collective provision. The distinction between any economy of scale which can be obtained only if the customers are located next door to each other and an economy of scale which can be obtained without this type of contiguity is fundamental.

AN ECONOMIC THEORY OF SOCIAL MOVEMENTS*

By ALBERT BRETON, *London School of Economics*
and
RAYMOND BRETON, *Johns Hopkins University*

Introduction

Social movements do not make their appearance at random; instead they tend to emerge in relatively well-defined circumstances and to be accompanied by identifiable socioeconomic factors. Consequently, in the present paper, we develop a hypothesis about the origin of social movements based on these "facts." To that end, the paper is divided into three parts: the first elaborates a theory of the demand for social change or, more specifically, a theory about why the demand for change manifests itself in social movements. This first part, in turn, is subdivided into two sections examining: first, some structural components of the environment in which social movements originate and discussing, second, the definitions and assumptions of the proposed theory as well as its essential mechanisms. The second part introduces the concept of the supply of social movements and examines how this supply is related to the demand for social change.[1] This discussion, incidentally, sheds some new light on the question of social change itself. This is briefly discussed in Section III.

I. *The Demand for Social Change*

If a small number of individuals choose to participate in one or more activities of a social and political character, these decisions should probably be analyzed predominantly in terms of factors related to the persons making the decisions; if, however, the number of individuals making the same or similar decisions is large, the environment in which the persons operate should also be examined. Since the social and political activities catalogued as social movements usually involve a relatively large number of individuals, it follows from the strategy just outlined that we should examine the environment in which these activities take place as well as the individuals engaging in them. It turns out that we can specify an important structural character of the environment which is sufficient to induce individuals to participate in activities aimed at inducing social change. We examine it immediately.

The Environment. Given that the constituent components of the social structure are highly interdependent, it is usually possible to characterize the environment by one carefully selected index. In the present study we use the rate of change in the level of social (national and regional) income or product for that purpose.[2] We assume that when the rate of growth in social income falls from a previously sustained high rate of growth to a lower one by a "significant" magni-

---

* This paper (or some earlier version of it) has greatly benefited from the comments of Yoram Barzel, James Coleman, Harry Johnson, Douglass North, Maurice Pinard, Arthur Stinchcombe, Pierre Trudeau, Gordon Tullock, and Mason Wade. We thank them and also assure them that they stand exonerated of its shortcomings and of any charges that may be levied against it.

[1] For other approaches to the formulation of the theory of social movements, see N. J. Smelser, *Theory of Collective Behaviour* (Free Press of Glencoe, 1963), esp. Chaps. 1 to 3; R. Heberle, *Social Movements* (Appleton-Century-Crofts, 1951); W. Kornhauser, *The Politics of Mass Society* (Free Press, 1959) and J. C. Davies, "Towards a Theory of Revolution," *Amer. Soc. Rev.*, 1962, pp. 5–19.

[2] This is done to simplify and to give operational power to the theory. However, any factor or set of factors that created a "significant" difference between individual aspirations and "reality" would serve our purpose as well.

tude and for a "significant" period of time, the environment will act as a constraining factor on the plans and expectations of individuals. What will constitute a significant reduction in the rate of growth and what will be the critical length over which this fall will be maintained for the social environment to be constraining will certainly vary from place to place and from time to time, but one would expect that the reduction in the rate of growth for an economy growing at 4 percent per annum would have to be at least 2 percentage points, and the time period over which the reduction should persist would have to be of at least two or three years. These, however, are empirical questions that should finally be settled on empirical grounds.

The Hypothesis. Since we will not be preoccupied with aggregation problems, it is immaterial whether we conduct our analysis in terms of the behavior of one individual or in terms of that of a group; consequently we will speak indifferently of one or the other, assuming that it is always possible to sum individuals into groups whenever that is necessary.

Consider, therefore, a typical group of individuals: at any point in time they will hold more or less definite plans about the future profile of their income and career plans. They can imagine that the time-shape of these flows and paths is such that their income and career prospects will point upwards for ten or more years and then level off, or that they will keep on increasing until they retire, or that they will have any of a number of possible contours. We assume that these individuals have more or less definite ideas about the factors which determine the size and behavior of these paths and furthermore that they classify these factors into personal factors such as experience, education, ability, aggressiveness and into environmental factors such as technology, the level of aggregate demand, the state of the balance of payments, and other similar factors which they assume to be outside of their control.

Individuals will have a tendency to weight heavily the environmental factors when comparing themselves with more successful individuals and personal factors when the opposite comparison is made. For example, individuals who experience a certain success relative to others will usually make reference to virtues (such as love of risk, initiative, hard work, courage, perserverance) and to vices (such as laziness, indolence, lack of initiative, unimaginativeness) in explaining their "success" and the "failure" of others. On the other hand, individuals experiencing failure relative to others will usually explain it in terms of bad luck, fatalism, their unfair exploitation by others and other environmental factors.[8]

The three foregoing assumptions relate to the individuals themselves and to the judgments they make about the determinants of their income and career history. The following assumptions relate to conditions which are outside the individuals. To be specific, we now assume that there exists an "objective" relationship defined in terms of "real" factors and "real" forces which are taken to be the "true" determinants of the income streams and of the career paths of individuals. It is therefore possible, given this set of assumptions, to analyze the difference between the imagined and the real weights given to each factor in the determination of income and to ascertain the extent to which the individuals hold "correct" opinions about themselves and the state of the world.

Now a protracted reduction in the rate of growth of economic activity will, by definition, manifest itself through environmental factors and since it is reasonable to postulate, in line with the third assumption made above, that the changes in the contribution of personal factors to changes in the behavior of incomes and career paths are either zero or known to the individuals, it follows that changes in their situation will lead them to look to changes in the environment for an explanation of what is happening to them. The analysis of the precise factors which individuals will select from the set of all possible environmental factors to explain their situation will be postponed until Section II. Suffice it to stress at this point that they will be environmental factors and not personal factors.

---

[8] The particular specification of weights sketched in the text is essential for the empirical analysis of who joins and who does not join social movements.

To summarize, we have assumed the existence of individuals who hold opinions about the size and time-shape of their income streams and career paths and about the factors determining them. We have also assumed that these opinions can be correct or incorrect when appraised against an objective benchmark; if individuals find that their opinions are incorrect, they will change them, though this does not imply that those they will adopt will be correct ones: they may simply change the weights given to personal and environmental factors so as to minimize the damage to their self-image; the analysis of this process has been postponed until Section II. Let us now, as a preliminary to this analysis, examine the impact of the appearance of environmental constraints on the intensity with which individuals desire to alter the environmental factors (the elasticity of the demand for social change).

In addition to the prospects they hold about their income and career paths, individuals hold ideas or plans about the rate at which they will consume the amenities which their incomes and their careers make possible. These plans can have any time-shape whatsoever, but for the sake of simplicity we assume them to be horizontal; that is, we assume that individuals plan a constant stream of consumption instead of a variable or steadily increasing or decreasing one. Both the planned income and consumption streams can be given graphic representation in the following diagram, which for simplicity also assumes linear income prospects.[4]

The curves $Y(t)$ and $U(t)$ represent the income and the consumption streams, respectively. It should be observed that at t_0 (the origin) the individuals are accumulating debts at a rate equal to AB/t_0B per unit of time; at t_1 that rate is zero, so that the stock of debts accumulated at t_1 is equal to ABC. These debts are accumulated because according to plans made at t_0, the individuals will, given their prospective incomes, have accumulated at t_2 assets equal to CDE sufficient to repay their debts (ABC) and the interest on them. Now imagine that as a result of a protracted reduction in the rate of growth of economic activity, individuals revise their opinions about their future stream of income from $Y(t)$ to $Y'(t)$. In Figure 1 we assume that individuals revise their expectations at t_0; we do this for the sake of simplicity and convenience alone; the following argument would remain unaltered if we assumed the revision of expectations to take place at any time after t_0. A look at that figure would seem to warrant the conclusion that individuals would as before borrow at a rate of $A'B/t_0B$ per unit of time instead of AB/t_0B and have accumulated at $t_1 + dt$ a total of $A'BC'$ debts, which could be repaid later on. Such an argument, however, would be incorrect, because even if we could assume a perfect capital market, it is unlikely that individuals would be able to borrow $A'BC'$, simply because, for some reductions in expected income at least, the debts would be too large to be repaid over the lifetime of the income stream. Only if individuals had a life-stream of asset accumulation that was large enough to repay $A'BC'$ could they borrow the sufficient resources. In general, what will happen is that individuals will reduce their planned consumption expenditures to some level between C and F (at F, $A'B'F = ABC$); they could, of course, reduce their planned consumption below F so as not to have to borrow $A'B'F$, which, being equal to ABC, represents in terms of $Y'(t)$ $[<Y(t)]$ a heavier debt burden. If they choose to hold their consumption expenditures above F, they would have to increase the absolute and the relative size of their debts. Most of the complications of the real world—imperfections in the capital market, repayment of a fraction of one's debts very soon after borrowing, relatively short planning horizons—all reinforce the effects of the mechanism just described.

---

[4] The argument in the text would only be changed in an inessential way if the income streams were $Y(t)^*$ or $Y(t)^{**}$ as may be more realistic to assume.

We reach the conclusion that as a result of the reduction in the rate of growth of social income, individuals are forced to readjust their own expected income prospects and this leads to higher debts and/or to lower consumption, both of these things in turn lead to a reduction in welfare, and on the assumptions stated above, this reduced welfare will be imputed to environmental factors. We further assume that the more important the required adjustment, the stronger will be the desire to transform the environment. This desire to change the environment in a still unspecified fashion we label the demand for social change.

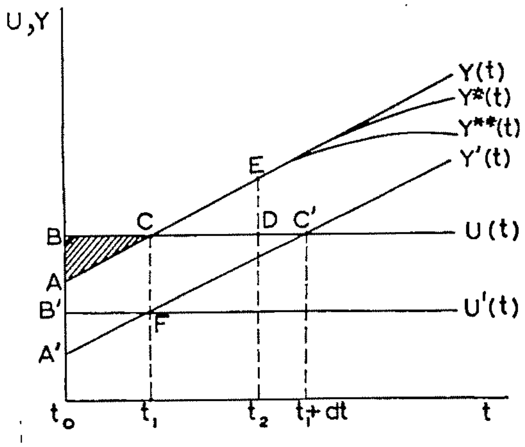


FIGURE 1

II. *The Supply of Social Movements*

In this part we start by assuming, in strict analogy with the economic theory of the firm and of markets which asserts that the demand for products creates opportunities for profits which entrepreneurs, after a more or less prolonged period of time, will want to make the best of, that the demand for social change creates opportunities for social profit which social entrepreneurs will want to reap and in so doing will "supply" or provide social movements to those who want them. Before examining this hypothesis any further we must pause to define the concepts of social profits and of social entrepreneurs which have just been introduced.

Social profits are in the nature of a flow of rewards which accrue to the social entrepreneurs and to their partners as long as the "product" which they sell is purchased. Social profits have both monetary and nonmonetary dimensions: the monetary dimension can be measured in terms of a flow of command over real resources which is made available to the entrepreneurs as a result of their activities. When social movements have reached their peak, this monetary flow of rewards can be sizable and for a given social entrepreneur often much higher than his alternative worth. This excess may be a quasi-rent to some specific quality which he possesses, but it may also be a transitory monopoly profit reflecting the disequilibrium of the market. If the later explanation is correct, the appearance of new social entrepreneurs entering as competitors will wipe the profits away; if the first explanation is valid, that is, the flow of monetary rewards is in the nature of a quasi-rent, we should expect some new entrepreneurs to be unsuccessful in their effort to attract the profits to them. The nonmonetary dimension of social profits is more in the nature of prestige and power. The efforts of social entrepreneurs to get their names in the mass media is an indication that for many of them this dimension of profits is a very important one.

We have also used the concept of social entrepreneurs. We use it in analogy with that of industrial and commercial entrepreneurs—concepts which can be found in economic theory and in the literature of the role of the entrepreneurs in economic growth. While the latter are busy promoting and developing industrial and commercial enterprises, the former devote their energies to social and political endeavors. At any point in time a society has a total stock of actual and potential entrepreneurs of the three types, who may be more or less good substitutes for each other. The allocation of this stock between the three types will depend first on the profitability of alternative occupations, a dimension which will depend in turn on the distribution of the demands of society for the products and services of the various entrepreneurs; if the three types of entrepreneurs are not perfect substitutes for one another, the distribution will also depend on their comparative advantages in meeting these demands. This allocation will also depend on the relative level of communication, information, and especially organizational costs; if the relative costs of organizing a social movement are *ceteris paribus* lower than the costs of promoting a business, more entrepreneurs will engage in the first of these two activities.

Let us suppose that as a result of a reduction in the rate of growth of social income some individuals in society have begun to revise the image of their environment and of the factors responsible for the alteration in their income, consumption, portfolio and career plans; they are, as it were, groping for a new *weltanschauung* or a new ideology.[5] It is the social entrepreneurs' role to furnish them with one, which should be simple enough to be capable of wide diffusion and still complex enough to have the appearance of a *weltanschauung* or ideology.[6] These will therefore generally be grounded in such difficult areas as money, race, morality, the place of the individual and/or of the government in society, and others of this nature. The social entrepreneur electing to emphasize the place and role of money in the economy, for example, will link all the troubles of the world to the behavior of the central bank, or to the stock of money, or to the level of interest rates. He will be identified with such movements as Social Credit and Thirty-Dollar every Tuesday, while the social entrepreneur who selects race as the factor responsible for the consequences of the constraining environment will be an antisemitic, or anti-American, or anti-Negro, and try to show that these groups are responsible for the difficulties felt by individuals. Another will blame the communists for all the troubles and difficulties of individuals, and he will be identified with the John Birch Society, for example; another one, still, will set out to show that big business is the culprit and he will sell a socialist ideology or *weltanschauung*. Depending on his background, his aptitudes, his knowledge and a host of other factors, lumped under the heading of comparative advantage, each entrepreneur will elect one factor (or maybe more than one for a new synthesis!) which he believes or at least wants others to believe is the prime cause of the difficulties through which some members of society are passing. The success of each social entrepreneur in selling his particular product will depend on the response which it receives; this in turn will depend on whether the individuals who are assumed to be in the process of changing the weights they had given to the personal and to the environmental factors in explaining their own altered personal situation elect to listen to one entrepreneur more than to another and to adopt a particular point of view. This choice in turn will be conditioned by the desire of individuals to minimize the damage to their self-image or self-esteem.

---

[5] This term is used in the precise sense given it by A. Downs, *An Economic Theory of Democracy* (Harper & Row, 1957).

[6] Social entrepreneurs sell *weltanschauungs*; in this way they resemble entrepreneurs selling advertisements, so that the characteristics of the *weltanschauungs* should in many ways resemble advertising messages. For a description and analysis of the latter, see H. G. Johnson, "Advertising in Today's Economy," in *The Canadian Quandary* (McGraw-Hill, 1963), pp. 269–84.

Before examining the types of behavior supplied by social entrepreneurs in addition to ideologies, we would like to state a number of testable implications of the above hypothesis which are either implied or contained in the foregoing discussion. A first such implication is that all social movements are accompanied by an ideology; that is, by a simplified (if not simplistic) explanation of the sources of trouble and difficulty facing individuals as a result of a reduction in the rate of growth of social income. A second, and one that should be obvious from the discussion of Section I, is that individuals with a high ratio of debt to income should have a higher propensity, *ceteris paribus,* to join social movements than others. A third is that a social movement will generally not appear in isolation; instead many movements will be observable at the same time. An exception to this could occur in the case of small reductions in the rate of growth of income generating only a weak demand for social change and in consequence only a small social profit capable of attracting and sustaining only one or very few social entrepreneurs. When the reductions are pronounced, many social entrepreneurs will wish to partake in the profits, so that many social movements will be supplied. Since in many instances small reductions in the rate of growth may not give rise to social movements of any significance, we would expect our general proposition to hold most of the time. A fourth implication is that the entrepreneurs leading the various social movements will be in strong competition with one another even when they are advocating substantially the same thing, and they will, therefore, spend considerable time arguing that other social entrepreneurs are wrong. This phenomeon, of course, will be much stronger when social movements are on the wane. Last, social entrepreneurs will devote a large part of their efforts in "differentiating their product" from that of competitors.

III. *The Process of Social Change*

In the introduction we mentioned that the theory of social movements developed in this paper had implications for the question of social change. We discuss this question here because it is an essential part of the theory outlined above.

The demand for change can manifest itself in at least three different ways: First, through social action of a magical character. In such instances, environmental changes are attributed to the actions of a supernatural being who must be persuaded or influenced to change the course of events. Second, through economic and/or political action by means of organizations and procedures set up for the purpose of instigating certain specific changes or change in general. And, third, through action of an economic and/or political character, but within the existing institutional structures, rules and procedures. This is essentially a mechanism characterizing democracy. The first two expressions of demand for social change are usually referred to as social movements, but it is obvious that the third expression could also be called a social movement.

Whatever kind of change is in demand, there are likely to be social entrepreneurs who will attempt to satisfy this demand, as we have argued above. The next question pertains to the conditions under which one type of action rather than another will be provided. When will social entrepreneurs be the leaders of existing institutions and when will they be initiating new associations or attempting to create new organizations? There are two ways in which the manifestations of the demand for change differ which are important in relation to this question. First, the location of the sources of control over the environment differs and, second, the extent to which this source of control is deemed manipulable also differs. A supernatural being can be seen as intractable or as responsive to expressions of deference and supplication of various sorts. Similarly, if the control over a situation is considered as being in the hands of certain individuals or groups, then the possibility of influencing these powerful individuals or of penetrating their organizational structure will give rise to actions of a representative nature: support for a given political party or for a certain party platform, pressure on labor, government, business or religious leaders. But if people find their leaders and institutions unresponsive to their problems, they will follow the leadership that develops marginally to the existing institutions.[7] They will support the new party that a political entrepreneur is creating; they will support political reform associations whose purpose it is to shake the government out of its lethargy; they will join manifestations of all sorts. Similarly, there will be social actions aimed at the social groups identified—rightly or wrongly—as being in control of the unresponsive institutions.

We are not advancing the scapegoat interpretation of social movements. On the contrary, we are arguing that people who are under the constraints of a protracted growth rate and who impute the cause of this to environmental factors will try to hit at those groups and institutions which are seen as either responsible for the situation or as responsible for doing something about it or both, and new movements will be supported when existing structures remain apathetic or, of course, hostile. For instance, a new political party is not likely to appear if the leadership of at least one of the existing parties is not tightly controlled by "conservative" groups (i.e., by groups unresponsive to the demand for change), thus allowing new issues and new policies to be introduced into the party platform. However, there may be groups who become the targets of social dissatisfaction but who cannot really do anything about it since they do not have any institutional power. Ethnic minorities and immigrant groups are frequently in that situation. Because of this, they appear as being treated as scapegoats, but in reality there is not really a "transfer" of aggressive tendencies on them; they are seen as part of the cause of the state of affairs, but they are not in a position to bring about a solution. In this sense, hostility against certain groups may be rational, but it is unlikely to be productive of any socially profitable change.

---

[7] In "One-Party Dominance and Third Parties," *Canadian J. of Econ. and Polit. Sci.*, Aug., 1967, M. Pinard has presented a complete model related to this point.

Because of the speed at which social and political institutions function, the reduction in the rate of income growth is likely to witness a number of social movements, even if political leaders are very responsive. It takes time to bring about effective solutions, especially to long-neglected problems. This is particularly the case when problems have been neglected because the elite did not have the resources to cope with them or simply did not know what to do about them. During periods of protracted reduced growth, then, both institutionally instigated changes and social movements are likely to occur; but the amount and strength of the social movements depend largely upon the responsiveness of the political and social elite.

Conclusion

The hypothesis formulated above appears to be consistent, at least in the large, with what is known of the history of social movements in Canada, in the United States, and in Europe during the first part of the twentieth century.[8] In particular, it helps to understand the appearance of the civil rights movement in the United States, of separatism in French Canada, of aggressive anti-Americanism in English Canada and of Welsh and Scottish nationalism in the United Kingdom during the current[9] protracted reduction in the rate of growth of these economies. A more systematic and refined treatment of data will not, we hope, have to wait too long.

[8] This history has been briefly sketched for Canada by the writers in "Le Séparatisme," *Cité Libre*, Apr., 1962.

[9] The exact dates vary from country to country, but they are all rather easily identified.

LEWIS CARROLL AND THE THEORY OF GAMES

By Duncan Black
University of Wales and *University of Chicago*

Lewis Carroll's booklet, *The Principles of Parliamentary Representation* (1st ed., Nov., 1884, 2nd ed., Jan., 1885), applies the technique of the two-person zero-sum game, which we usually associate with economic theory, to provide a theory of proportional representation and a theory of the apportionment of parliamentary seats. The entire reasoning of the booklet is expressed in quantitative terms, again on the basis of the two-person zero-sum game and the maximin criterion. A few months after the booklet had been completed Carroll made a further application of game theory, this time using the coalition game, in a very tentative way, to find the most suitable set of candidates to represent a constituency in an election.

The year 1884 was the year of the last great debate in Britain on the franchise. If we exclude Ireland, the electors belonged to one or other of the two political parties. The number of parliamentary seats was fixed, and the more seats the one party got, the fewer went to the other. This setup invites an application of the two-person zero-sum game; but to the present day, with a solitary exception,[1] Carroll has been the only writer to provide a model of this kind. Whether we say that he actually uses the two-person zero-sum game in which each of the parties acts on the maximin criterion, or on the other hand provides only a model which has the same effect as this, is a semantic question and perhaps not very important. In any case, my own answer would be that it is the two-person zero-sum game and the maximin criterion which he uses; and there can be no doubt whatever that he makes use of the coalition game.

Carroll regards the choice of an electoral system as being a problem in proportional representation. Consider, he says, the main family of electoral systems: that in which a constituency has m seats ($m \geq 1$), and the elector is allowed v votes ($1 \leq v \leq m$), of which he can give no more than one to any candidate.[2] The seats will of course be awarded to the m candidates with the highest number of votes. The best or most suitable electoral system will be that member of the family which gives rise to the greatest degree of representation. Or, in his treatment, going on to quantify the problem, the best electoral system will be that member of the family which maximizes "the number of electors represented," and which consequently minimizes "the number of electors unrepresented." This requires, among other things, a definition of "the number of electors represented" (or unrepresented) in any particular election; but we will come to that later.

To get his model of the two-party system in politics, Carroll must, of course, make certain definite assumptions. He supposes that each of the two political parties knows the number of its own supporters and the number of supporters of the other party; and that each party is able to direct the voting of its supporters as it chooses, getting so many of its supporters to vote for these candidates, so many for these others, and so on, with a view to maximizing the number of seats it fills. To illustrate, let us choose as a particular example a constituency contested by two parties, A with 100 supporters and B with 70; and let us take it that this constituency has three seats ($m=3$) and the elector is allowed two votes ($v=2$). The party A may choose to put up two candidates giving each 100 votes, i.e., may use the strategy (100, 100); or it may put up three candidates and use the strategy (100, 60, 40), or the strategy (80, 70, 50), and so on.

---

[1] Cf. Jack Sawyer and Duncan MacRae, Jr., "Game Theory and Cumulative Voting in Illinois," *Amer. Polit. Sci. Rev.*, 1962, pp. 937–40.

[2] The booklet deals also with cumulative voting, in which the elector may cumulate his v votes on one or more of the candidates.

To tackle this problem of the three-seat two-vote constituency, we might, at the present day, argue in this fashion. The set of strategies open to the party A is the set of ways of partitioning 200 votes into two parts each consisting of 100, together with the set of ways of partitioning 200 votes into three parts no one of which exceeds 100. And conceptually we may arrange this rather long series of strategies down the left-hand column of the payoff matrix. Likewise along the top row of the matrix we may arrange all the strategies open to the party B.

Now with a matrix framed in this way, suppose that A uses any given strategy open to it and B uses any given strategy. The electoral rules will then specify which of the candidates are to get the seats, and we can fill in the figure in the corresponding cell of the matrix, showing A's payoff and B's payoff. When we fill in each cell in this way, we can proceed in the usual manner by adding a column of figures to the right of the matrix to show the minimum number of seats A fills, whichever strategy it may use, and correspondingly for B. From this, if we suppose that each of the two parties acts on the maximin criterion, we can deduce the choices made by both A and B. Approaching the problem in this way we might be apprehensive that choice by each party on the maximin criterion might not give rise to an equilibrium point—and in this problem the concept of a "mixed strategy" would be meaningless. It can be shown, however, that in the electoral problem we need have no cause for apprehension and that, trivial exceptions apart, choice made on the maximin criterion must give rise to equilibrium.

This shows the logical structure of the problem and it is the way in which we might approach it today; but in Carroll's day a different approach appears to have been quite common. In 1884 almost two-thirds of the parliamentary seats were in multi-seat constituencies returning two, three, or four members, in which voting tactics were vitally important and came under supervision of the local party caucus. Already in 1853 in a pamphlet[3] eulogized by John Stuart Mill, James Garth Marshall had shown how, in what we now refer to as the m-seat v-vote constituency, each of the two parties may choose a rational strategy. Marshall had worked out a large number of arithmetical examples and, although he did not formulate these rules, it was made fairly plain that to choose a rational (maximin) strategy it was sufficient that a party should: (1) aim to fill a definite number of seats, s seats say ($v \leqq s \leqq m$); (2) put up exactly s candidates; and (3) distribute its votes among its s candidates as evenly as possible.

This provided a very direct way of choosing a rational strategy. Knowing its own strength and that of its opponent, a party needed to consider the outcome only if it put up v or $(v+1)$ or . . . or m candidates, dividing its votes among them as evenly as possible, while its opponent, behaving rationally, did the same from its side. Take, for instance, the three-seat two-vote constituency with 100 voters of whom x support the party A and $(100-x)$ support the party B. A rational strategy for A would be either

$$(x,\ x) \text{ or } \left(\frac{2x}{3},\ \frac{2x}{3},\ \frac{2x}{3}\right)$$

---

[3] *Majorities and Minorities: Their Relative Rights.*

—though other rational strategies may exist which have the same outcome as one of these; and a rational strategy for B would be either

$$(100 - x, 100 - x) \quad \text{or} \quad \left(\frac{200 - 2x}{3}, \frac{200 - 2x}{3}, \frac{200 - 2x}{3}\right).$$

Suppose x is known and equal to 46. A rational strategy for A would be (46, 46) or (31, 31, 30), "or" being used in the exclusive sense, and for B (54, 54) or (36, 36, 36). If A puts up two candidates with 46 votes apiece, it is bound to fill one seat against the best counterstrategy that B can bring against it, though the stategy (44, 42) say, would also secure this. Similarly a rational strategy for B to use is (54, 54), ensuring that it fills two seats against whichever strategy A may employ; and no strategy exists ensuring that B will fill more than two seats.

Marshall's own work was cast in terms of critical values—the minimum number of supporters required by a party to fill one or two or . . . or m seats, in some arithmetical example. From the above it is easy to verify that, in the three-seat two-vote constituency, to fill one seat against the best counterstrategy, a party requires the support of 40+ p.c. of the electorate, to ensure filling two seats the support of 50+ p.c. and to ensure filling three seats the support of 60+ p.c. In terms of these critical values, a party that has the support of 46 p.c. of the electorate will again be able to ensure filling one seat but not more.

Marshall's reasoning implies acceptance of what we know today as the maximin criterion in the two-person zero-sum game, and the first step in Carroll's argument is to generalize this type of reasoning for the m-seat v-vote constituency. Carroll shows algebraically the volume of support sufficient to ensure that a party which acts on the maximin criterion will fill one or two or . . . or m seats, and throughout the entire chain of reasoning of the booklet the implicit assumption is that both parties act on the maximin criterion.

The next step in the argument derives from game theory and, as we would formulate it today, makes explicit use of the maximin criterion. The purpose of the booklet is to show which constituency, i.e., which values of m and v, give the maximum degree of representation. Applying Occam's razor, Carroll avoids definition of "representation," a complex concept which he does not require, and defines only "the number of electors represented" or equally its converse, "the number of electors unrepresented," for if, out of the e electors in a constituency f are represented, $(e-f)$ will be unrepresented.

His definition of the percentage of electors unrepresented is this. Suppose that in any given election in which both parties act on the maximin criterion, a party with the support of h p.c. of the electors fills a given number of seats, but could have filled this same number of seats with the support of only h' p.c. of the electorate $(h' \leq h)$: then the votes of $(h-h')$ p.c. of the electorate who support this party are useless or "wasted" and play no part in determining the result, and $(h-h')$ p.c. of the electorate is unrepresented. Likewise, if the other party has the support of k p.c. of the electors and could have filled the same number of seats with the support of only k' p.c. $(k' \leq k)$, then a further $(k-k')$ p.c. of the electors are unrepresented. In all $(h-h')+(k-k')$ p.c. of the electorate is unrepresented.

To illustrate, take again the three-seat two-vote constituency in which one party has the support of say, 46 p.c. of the voters and, with both parties acting on the maximin criterion, fills one seat. But it could have filled one seat with the support of only 40+ p.c. of the voters. Hence 6− p.c. of the voters are unrepresented. Similarly the other party with the support of 54 p.c. of the voters fills two seats and could have done so with the support of only 50+ p.c., so that a further 4− p.c. of the voters are unrepresented. In all in the election 10− p.c. of the voters are unrepresented.

From these two premises, the argument proceeds to establish the mathematical expectation of the percentage of the electorate unrepresented in constituencies with various numbers of seats and with the elector allowed more or fewer votes. It is found[4] that in the m-seat v-vote constituency the mathematical expectation of the percentage of the electorate unrepresented diminishes as m increases and again is smaller the smaller the size of v. The commonsense conclusion to be drawn is that we want an electoral system based on large constituencies, with four or five seats, Carroll suggests, in which the elector is allowed only a single vote.

The model for the two-party system is got on the basis of various simplifications, one of them being the assumption that the elector prefers any candidate of his own party to any candidate of the other, and another that he is equally well represented whichever candidate or candidates his party chooses to put up. But before he wrote the booklet, in letters to the *St. James's Gazette,* and after he had got it completed, in a *Supplement* and *Postscript to Supplement,* Carroll employed a more general model than this. Suppose we have a multimember constituency in which, in relation to the candidates who stand, the voter's preferences are subject to no restriction, and possibly do not even follow party lines. How, with m seats available, can we choose that set of m candidates which will give rise to the fullest representation?

Carroll gives what we may term (1) a conceptual approach, (2) an operational approach, and (3) a practical scheme of election. In all three he employs the Droop quota: in this m-member constituency it is possible (in all relevant cases) for each of exactly m candidates to get a full Droop quota of votes.[5]

1. In the conceptual approach we may imagine a group of voters, each with a definite preference schedule in relation to all the candidates who stand, and each voter knowing the preference schedule of every other. We would allow the electors to form and reform themselves into coalitions, each coalition aiming to command one or two or three, etc., Droop quotas of votes, in order to fill accordingly one or two or three, etc., seats. If such coalitions were able to form and re-form by a process of contract and recontract, until a stable set of coalitions emerged from which no elector had any incentive to detach himself, we would regard the candidates returned by this set of coalitions as an optimum set of representatives for the constituency.

This would be possible in practice, however, only if (and of course here and elsewhere we are using modern terms[6] where they help to express Carroll's notions) the costs of obtaining information about the preference schedules of other voters, and about the state of the coalitions at any moment, were zero or very low, and if the costs of entering into fresh contracts with the members of any coalition were zero or were very low, and if the process could be carried out within a short period of time. In fact, except for the very small group, the costs of the elector finding out the preferences of other electors, finding out the existing state of the rival coalitions, and of terminating the existing contract and entering into a new contract with another coalition would be far from zero. It is just not feasible in the ordinary election for the voters to form and re-form themselves into coalitions, one set after another, until a stable set is arrived at which will return m candidates.

---

[4] Cf. Black, "The Central Argument in Lewis Carroll's *The Principles of Parliamentary Representation,*" *Papers in Non-market Decision-making,* No. 3, Fall, 1966, pp. 1–17.

[5] The Droop quota is the integer next greater than $\dfrac{\text{total number of votes cast}}{m+1}$.

[6] Cf. James M. Buchanan and Gordon Tullock, *The Calculus of Consent* (1962), Chap. 8.

2. We may therefore seek an operational approach. Take it that we have a given set of preference schedules. The voters cannot in practice enter into coalitions among themselves, but we may be able to draw up rules whereby we can discern the coalitions which the voters, acting rationally, might form. This would show which individual candidates would get the support of 1-Droop-quota coalitions to elect these individual candidates, which pairs of candidates would get the support of 2-Droop-quota coalitions to elect these pairs of candidates, and so on.

Carroll in fact sketched a few rules which would assist in picking out coalitions of this sort, but said they were incomplete; and after working longer on the problem he probably got to know that, in general, the answer to the problem is indeterminate. But, in particular examples, determination of the coalition which it would be appropriate for the voters to form seems clear enough to common sense, and Carroll gives two or three instances of this sort: in fact he employs them to show that the single transferable vote gives a different and therefore a wrong answer in these particular cases. In general, however, an operational answer to the problem is again not feasible.

3. In the practical scheme of election which he proposes for use in Great Britain, Carroll sidesteps both the previous difficulties. He argues that, in the first place, the average British elector, "Hodge, fresh from the plough!" will know which candidate he prefers to any of the others, but will be unable to rank the candidates in order of preference: his felt preferences will amount to him knowing only which of the candidates he likes most. If so, the elector should be asked to give a vote for only a single candidate. Then at the end of the election the candidates themselves, Carroll suggests, should collect up their votes and treat these votes as if they were their own private property. Any candidate with a Droop quota of votes should be elected. When this had been done the next stage would be for the candidates to meet and exchange their votes among themselves, those who had already been elected exchanging or donating their surplus votes over and above the Droop quota which had been used in their election. It would be at this stage in the process that the coalitions would be formed. The candidates would already know one another's political attitudes and, with only a few well-informed and well-practised people meeting together after an election, the transaction costs in the formation of coalitions would be very low. Thus in casting only first-preference ballots the voters would have expressed accurately their attitudes; and, at the same time, the choice of an optimum set of m candidates as envisaged by the conceptual approach, would have been attained by the actual formation of coalitions—but coalitions among the candidates and not among the voters; i.e., among people to whom coalition formation is appropriate.

We will not attempt to evaluate this suggestion of Carroll's. For our present purposes we wish only to point out that just as in the two-party contest he makes use of the two-person zero-sum game, so in the nonparty or multiparty case Carroll makes use of the basic notions of what we know today as the coalition game with ordinal utilities.

DISCUSSION

Thomas E. Borcherding: Professor Black's scholarship has unveiled another interesting and important contribution of Lewis Carroll to the theory of collective choice.[1] Carroll, we now discover, beside his interest in cyclical majorities, was keenly interested in devising an ideal method for the democratic representation of voting constituencies in legislatures and parliaments. The task would be difficult enough today, but given that the first edition of Carroll's booklet is dated 1884, I am indeed impressed. Intriguing as Carroll's exploration is, I find that his actual contribution to our current understanding of the problem is not as fruitful as I had hoped. Still though I find Carroll's model inadequate, I am far from disappointed, since the issue to which he addresses himself is most important.

It is my purpose to scrutinize the conceptual foundations of Carroll's electoral *desideratum* which I have named, for want of a better term, the "Representation Optimum" (RO). This principle states that given that an electoral district has a fairly large number of voters, it should be represented in the legislature by several seats (perhaps four or five would be sufficient) and each voter should be permitted a single vote.

The RO depends crucially on at least two things: first, that in the real world where electoral contests can be viewed as two-person zero-sum games, rational players always act on the maximin and minimax criterion, and, second, that Carroll's definition of nonrepresentation is reasonable.

I will accept the characterization of the two-party electoral process as a zero-sum game. I wonder, however, if we should consider the maximin as the only criterion that a rational, i.e., utility maximizing, party would choose? Does utility maximization in the Carroll-Black description of the game necessarily imply determinate outcomes? My answer to both these questions would be no.

Ellsberg pointed out several years ago,[2] and rightly I believe, that the conservative play dictated by the maximin criterion, while certainly holding for the minorant and majorant versions of two-person zero-sum games with saddle points, does not necessarily obtain in the normalized version of that same game. In the minorant version one party (call it A) makes the initial move and the other party (B) then follows with its choice. Since interests are diametrically opposed any choice of strategy by A other than the minimax will surely be irrational. A chooses his strategy knowing with certainty that B will then pick a strategy that will do A the most damage and make B as well off as possible. In the majorant version the order of play between A and B is reversed and B will initially (and rationally) choose his maximin.

The normalized version of this game is quite another thing again, even given that both A and B face the same imputation matrices. True uncertainty enters since choice is simultaneous and, hence, neither party knows the opposition's strategy until afterwards.

If the payoff matrices to A and B are objectively given rather than subjectively estimated and if each choosing party's estimation of the probability of error by the opposition is taken to be very slight or even zero, the result is determinant and identical to the minorant and majorant versions. However, if we relax these assumptions, as I think we must for application to the real world, it becomes conceivable that the expected payoffs from choosing a non-minimax (or non-maximin) strategy may at times be larger than the expectations of returns under the conservative play dictated by the minimax or maximin.

This becomes even more apparent when we realize that in the real world there often exists sizable numbers of uncommitted voters, that the elasticity of voter support with respect to party platforms is not zero,[3] and information about voters' preferences is costly to obtain. In these circumstances A's and B's assessments of their respective payoff matrices will differ both between themselves and with the matrix that contains the imputations that actually exist. Further, even if A and B should be very accurate in their assessments, they know this only *ex post*.

---

[1] Carroll's understanding of cyclical majorities was discussed in Black's *Theory of Committees and Elections* (Cambridge Univ. Press, 1957), Chap. XX.

[2] Daniel Ellsberg, "The Theory of the Reluctant Duelist," *A.E.R.*, Dec., 1956, pp. 909–23.

[3] The term "party platform" implies all forms of

It is for this reason that I believe that we cannot hypothesize a priori the employment of the maximin-minimax criterion by the choosing parties. This, in turn, implies an indeterminacy in the outcome of play and, hence, an inability to predict accurately, using Carroll's definition, the extent of nonrepresentation.

The situation certainly becomes even more unpredictable when voter preferences are distributed polymodally as at times they well may be. Under such a preference distribution three or more parties may develop as Downs has suggested.[4] Although neither Carroll nor Black develop a model for three-party contests, I would venture that my doubts on determinacy would, under such circumstance, be greatly magnified.

But let us assume that we found by subjecting electoral data to empirical tests that the maximin criterion was the only rational course of play followed. I still am unwilling to accept completely Carroll's RO because I fundamentally disagree with Carroll's definition of nonrepresentation. I find Carroll's definition wholly at variance with the definition ordinarily employed or implied by choice theorists. For instance, I find that Tullock's conception of representation is more typical and reasonable. He views a representative as an agent of the voter who mirrors the preference structure of the individuals whom he represents.[5]

Carroll's definition is not based on individual preference functions but is purely mechanical. To give an example, suppose that a district has three seats, 100 voters, each with a single vote, and two parties, A and B, who receive 45 votes and 55 votes, respectively. Carroll would have us believe that 8 voters are unrepresented since A could have won its single seat with only 41 votes and B, its two seats with 51 votes. This holds whether or not A and B provide representatives that are perfect agents for their respective supporters. Such a definition of representation seems to have little more to recommend it than convenience of computation.

Assuming that we now possess an adequate theory of elections, what we need to find for the RO is a function that somehow measures in aggregate the costs and benefits of each proposed system (m and v in Black's terms). In the real world where transaction costs are not zero, m will surely be far less than the number of voters since decision costs rise both with the number of representatives and the number of voters in the constituency. To circumvent the aggregation problem we might employ the device used by James Buchanan and Tullock in *Calculus of Consent*.[6] They assume a representative chooser who is uncertain of his exact economic and social position in the future but who has a notion of the distribution of these states. Since this distribution is assumed to hold for each citizen, the choice of a RO by one will hold for all since it represents in an expectational sense that institution which has the highest discounted payoff.

A. ALLAN SCHMID: Issues of group and individual choice are steadily being illuminated and raised from unexamined habit and dogma to a subject for reason. The three authors in this session have made important interrelated contributions. Tullock discusses the characteristics of products whose provision collectively seems more efficient than in private markets. Black discusses the rules for selection of those representatives that will make detailed public choices and the Bretons outline a theory of how demand for group goods arises and gets transmitted.

Tullock reviews the special sort of "economies of scale" which arise from the technolog-

---

persuasion and action that form expectations in voters' minds about the bundle of goods and services that will be offered should the party's candidate be elected.

[4] Anthony Downs, *An Economic Theory of Democracy* (Harper and Row, 1957), pp. 122–27.

[5] Gordon Tullock, *Toward a Mathematics of Politics* (Univ. of Michigan Press, 1967), Chap. X.

[6] Univ. of Michigan, 1962.

ical characteristics of certain goods and which make their cost cheaper through group purchase rather than individual. These gains are offset by certain other decision and bargaining costs and costs arising from the fact that all must share taxes for group purchases but not all have identical tastes. Once the decision is made to have a group purchase, this still leaves the question of whether each expenditure is to be voted on separately or in clusters. The more expenditures are clustered and the larger the group making the decision, the more likely that member tastes will vary and noise problems arise in communication of these tastes. Yet, we find that certain things must be done by a large group such as those represented by the U.S. Congress.

Is there anything that can be done to reduce these noise and decision costs? The alternative is not only individual referenda or legislative bodies. The individual referenda and other communication with the legislature might be reduced if certain broad criteria could be discussed. A voter has a chance to understand and express himself on the criteria for whole varieties of programs but would be lost if he had to communicate on each separate expenditure.

At the risk of opening an area which is not so neutral as mosquito control, consider such things as the River and Harbor Omnibus bill. Currently the congressional debate and constituent communication is around a list of many individual projects. This is a type of cluster. Debate on dozens of bills seems very expensive and so we turn to an omnibus bill. Yet, the noise problem is large as it is difficult for voters and congressmen to get enough information to evaluate the package without considerable cost. What would be the economics of Congress voting on broad criteria for the program which the administrative agencies could then use to select individual projects? As already mentioned, the voter cost of getting the information necessary to decide on the broad direction and to transmit it may be less. What are the offsetting cost increases? Those to the voter directly may be small. One dimension is that decisions would be less incremental. It might be harder to experiment with a new direction without committing the whole program. But even if the voter could reduce his net direct costs, action in the form of broad program criteria may create costs for legislators. The cost of getting a compromise on criteria may be much higher than on decisions involving project additions or deletions to the President's budget list. Broad criteria raise ideological issues which are avoided in mere lists of expenditures. It is conceivable that the legislative decision costs may be prohibitive. Much ingenuity will be needed in testing the costs and gains of congressional voting on program criteria versus lists of individual projects.

The question of clustering, as a subject for research, is closely related to that of sequencing. The legislative success of a certain cluster of expenditures may depend on the cluster of projects that preceded it and the cluster anticipated to follow. The attitude of a legislator and his constituents toward a certain bill depends on whether they have already voted on other bills that are of particular interest and what promises for future support had to be made.

Rules governing the sequence of votes on spending proposals in legislative bodies are few and mostly unstated. Such bodies usually have procedures for setting the calendar and determining when bills reach the floor for debate, but the rules and criteria for this are seldom formal and the subject of public debate.

The advent of multiyear program planning and budgeting analysis in the federal government raises some important issues for the sequence of votes on expenditures. Now, projects in a number of program areas are voted on as they become available with little explicit attention to whether a better project is coming up in the future. The projects in a particular program are voted on in one-year clusters determined by current availability. If the future supply of projects were known, for example, in five-year clusters, the debate could consider the trade-offs between projects of present and future availability. This public choice rule change would produce a different ball game. While there would be gains in the scope of alternatives that are explicitly considered, it has costs in terms of data inaccuracy and loss of political flexibility. Nevertheless, the

1967 *Report of the President's Commission on Budget Concepts* found the case for public disclosure of longer-term budget projections very persuasive. Research on these issues should have high priority.

Tullock, like all good welfare theorists, reaches the issue that while total gains may exceed costs, there may be problems of distribution. We assuage our consciences by hoping that the gains and losses are random over a variety of decisions and time. Yet, one suspects that some groups consistently help pay for certain goods that they do not benefit from and receive no offset from other group goods. Tullock acknowledges that clustering of decisions may allow logrolling to reflect the relative intensity of different demands. Is it possible that certain government programs are or could be used in large part to compensate certain groups for losses they suffer when the majority rules on certain other expenditures they do not want? For example, an Indian lands program could be used to reimburse the mosquito-proof Indian who is forced to pay for mosquito control. This makes program analysis very difficult for the analyst never knows what the full objective might be. I recall a Presidential staffer who prepared what he thought was a logical analysis of a certain agricultural bill only to find it ignored when the President traded it for support for a civil rights bill. Is that an abomination or a reasonable procedure for compensating certain groups for their contribution to other group purchases which they do not want?

Tullock concludes that geographical contiguity is at the heart of an economics of group choice. This is a large part of it, but should not be carried too far. I am afraid we have seen the group too much in tribal terms where the tribe is defined in the immediately observable factors like geography, class, and language. However, the range of cost reduction from group choice is not alone a function of being "next door." Certain technologies make groups out of widely separated people.

While much useful work remains, I judge that, currently, the highest marginal returns are not in further knowledge of optimal public choice and constitutional rules but in how to institute those we now understand. The economics of institutional change has not yet had its Pigou and latter-day Buchanan's and Tullock's. I am constantly impressed by the failure of our institutions to move. What will it avail researchers to point out more optimal forms of institutions if we do not know how to get them instituted. Public announcement of research results is not enough. For example, I doubt if I shall ever have the audience of a *Life* magazine, but its exposure of a Mafia-connected New Jersey congressman did not prevent his recent reelection.

The Bretons develop a suggestive theory of social movements. Yet it largely says that if sufficient demand for change is created, politicians will try to supply it. This leads to the conclusion that the demand I feel is not shared by others, or at least if they feel it, it is not translated from personal misfortune to possible group actions. The economics of this translation needs attention. It has been said that to be a reformer is frustrating and to be reformed is maddening. Can economists reduce the costs of this frustration and madness?

BENJAMIN WARD: The Bretons have provided us with a sort of sore-loser theory of social movements. National or regional income growth slows down, thereby frustrating consumption and asset expectations. An "ability illusion" leads many to attribute a more than reasonable share of their problems to their environment rather than to personal inadequacies, thereby magnifying the demand for social as opposed to individual change. Social entrepreneurs, noticing the opportunity for personal gain, both monetary and nonmonetary, step into the breach, offering an ideological explanation for the situation which tends to exonerate an appropriate segment of the suffering multitude from any blame for their predicament, and at the same time provides an organization geared to bringing about some sort of redress. Because the frustrations are widespread, a number of such entrepreneurs make their appearance and keen competition for the sore-losers' allegiance sets in.

There is a good deal of plausibility to this thesis. For an American the natural testing ground would seem to be the depression years and here things seem to work out just as predicted. Technocrats, money cranks, racists, Communists, New Dealers, conservative Republicans and a variety of other "social movements" were competing under aggressive management, and with differentiated ideologies which exculpated the worthy, for the allegiance of those whose debts had been trending up while their incomes were trending down. Perhaps there is even discernible a tendency for many of the time-trend losers to be sore losers, though in such a strong case that assumption may not be necessary. All three of the types of proposed redress were in evidence, from the incantations of the millenarians through the revolutionaries to the ideologically-oriented politicians within the established parties.

However, despite this successful, if very superficial, test of the theory, there seem to be some serious problems with it, not all of which can be resolved by a more carefully qualified and extended exposition. One such problem is essentially semantic: just what do the terms "social movements" and "origin" refer to? As will be indicated, unless these terms are very carefully defined, it appears that there are important counterexamples to the theory. On the other hand, if they are too carefully defined, the conditions establishing applicability of the theory may become so complex as to destroy much of its explanatory usefulness.

There are three aspects of the theory which appear to be of less than universal applicability. The first of these has to do with the appropriateness of the single index, national or regional income, as the predictor of the occurrence of social movements. Often the frustrated citizens are those who are left behind, at least relatively, by economic progress, such as cotton or wheat farmers, or old folks, or ethnic minorities. In some cases these people are concentrated regionally, so that the regional income index picks up their plight. Perhaps this was true of the American Populists, who seem to have "originated" during a time of rising incomes, in terms both of national averages and even of farmers overall. But where the frustrated group is scattered over the country, as is the case with some ethnic groups who have contributed to the formation of social movements, these indices do not reflect their economic status. Of course, it is perfectly possible in principle to find a more appropriate index, but this raises the dilemma of optimal sharpness of definition mentioned above. The less aggregative and more complex the environmental index, the less powerful the theory.

The second problem also deals with the environmental index. Is it true that social movements only arise after a downward movement in an economic growth rate creates frustrations in a large number of people? The social movement led by George Wallace does not seem to have originated in a time of falling growth rates, either nationally or regionally, though there was certainly an economic component to the fears expressed by many Wallacites. That is, some social movements may arise because of the fears of some individuals regarding their future economic status, though the fears were not triggered by a decline in some economic index. Of course, the theory can be interpreted as providing sufficient but not necessary conditions for the origination of social movements and still retain considerable interest.

The third problem has to do with the role of economic self-interest as a trigger to movements and as a motivator of participants. Latter-day McCarthyism seems to have been triggered by a limited war during a prosperous time and to have attracted relatively large numbers of middle-class college students who were themselves exempt from participation in that war. McCarthyism seems to satisfy such descriptions as the authors have given as to the properties of social movements, including the desire for economic/political change, except for the question of economic interest. It appears that other than economic frustrations are an important part of some social movements, a seriously complicating factor, even for the sufficient-condition version of the theory.

Some years ago sociologists and historians were debating the precise meaning of the concept of a social movement. Schools formed around various definitions which themselves began to acquire some of the attributes of social movements—including, no doubt, an economic component. Some of these theorists put primary emphasis on the content of the ideology of alternative movements. The present authors have added a welcome counterweight to such extremism. But the substantive content of ideologies surely often has an important effect even on the origin of social movements. For example, a good deal of preparatory intellectual work may have to be done before a particular ideology is usable for political purposes; often, too, the environment must have rather specific properties for the potential movements to "originate." Though they have brought out some frequently very important factors determining the origins of social movements, the authors have not shown that other factors can safely be ignored.

THE EFFICIENCY OF EDUCATION IN ECONOMICS

THE EFFICIENCY OF PROGRAMMED LEARNING IN TEACHING ECONOMICS: THE RESULTS OF A NATIONWIDE EXPERIMENT*

By RICHARD E. ATTIYEH, *University of California, San Diego*, G. L. BACH, *Stanford University* *and* KEITH G. LUMSDEN, *Stanford University*

Summary

Recently considerable interest has developed in the use of programmed learning in the field of economics. While previous experiments[1] have been on a small scale, the results suggest that programmed texts can be very effective teaching materials. This paper reports the results of a larger, nationwide experiment, involving forty-eight schools and 4,121 students, designed to evaluate the efficiency of programmed materials in teaching the core micro- and macroeconomics sections of the typical elementary economics course. The major results of the study are:

1. On average, by spending twelve hours studying a programmed learning text students learned practically as much micro- or macroeconomics as did students in seven weeks of a conventionally taught elementary course.
2. On the basis of the test question breakdowns, students who used only programmed learning materials, as compared to conventionally taught students, performed better on "applications" of theory than on simple "concept recognition."
3. Students had a generally positive attitude toward programmed learning.

These findings suggest that the basic concepts and tools of micro- or macroeconomics can be self-taught in about two weeks' time with programmed learning materials, thereby freeing a much larger portion of the total course to help students develop skills in the application of the basic theory to "real world" problems.

Experimental Design

The principal objective of this experiment was to compare the performance of students using programmed learning, either by itself or as a supplement, with that of students taking a conventionally taught elementary course. Student performance was measured by test scores on an independently devised, nationally normed multiple choice examination taken when the micro- or macroeconomics topics were completed. To separate the effects of the programmed materials from other variables, information was also obtained on the educational level, sex, and scholastic aptitude of students, the type, size, and quality of schools attended, and the textbook, class size, and experience of teacher for conventional sections.

Each participating school established three test groups. Students in Group I were given copies of one of two programmed texts and were told to study this text exclusively (avoiding other texts and not coming to class) for a period of time determined independently by each school.[2]

---

* The authors are grateful to the Joint Council on Economic Education for financial support, to Prentice-Hall, Inc., and McGraw-Hill, Inc., for providing programmed materials at cost, and to participating colleges and universities for their unstinted cooperation.

[1] See K. G. Lumsden, "The Effectiveness of Programmed Learning in Elementary Economics," *A.E.R.*, May, 1967, and "Technological Change, Efficiency and Programming in Economics," in *New Developments in the Teaching of Economics*, ed. K. G. Lumsden (Prentice-Hall, Inc., 1967).

[2] The programmed texts used were R. C. Bingham, *Economic Concepts: A Programmed Approach* (McGraw-Hill, Inc., 1966) which contains both micro and macro sections; or either K. G. Lumsden, R. E. Attiyeh, and G. L. Bach, *Microeconomics: A Pro-*

On average, students in this test group had three weeks to read the programmed book, of which they used, according to their responses on a questionnaire, only twelve hours. At the end of this period they were tested, and then rejoined or became a conventional class.

In both Groups II and III, students were given conventional reading assignments and attended class lecture and discussion sessions. Students in Group II, however, were also required to read a programmed book, generally at a time and pace of their own choosing, while students in Group III were asked not to use the programmed book. In each school these two groups were tested at the same time, usually several weeks later than Group I. Each school had as much time as it wanted to cover the basic micro- or macroeconomic analysis. On average, students in these two groups had seven weeks to prepare for the examination and students in Group II spent eight hours reading the programmed book.

The tests used were preliminary forms (two micro and two macro) of the *Test of Understanding in College Economics* (TUCE) prepared at the suggestion of the A.E.A. Committee on Economic Education by a special committee sponsored by the Joint Council on Economic Education.[3] To avoid possible contamination of the results for Groups II and III from the earlier testing of Group I, the test form for these two groups was different from that given to Group I.

The sample was drawn to provide adequate coverage for five main types of colleges and universities—"high prestige" schools, liberal arts colleges, large universities, state colleges (often formerly called teachers colleges), and junior colleges.[4] Of the 4,121 students on whom we were able to acquire a complete file of data, 84 percent were male. The average student was a sophomore and had the equivalent of an SAT composite score of 1108. The average school had 10,620 students and an entering class with the equivalent of an SAT average composite score of 1122. The typical student in Groups II and III was in a class of seventy-seven, and had an instructor who had been teaching for 6.8 years. Sixty percent of the students in these two groups used one of three widely known textbooks, referred to as textbooks A, B, and C. The remaining 40 percent used one of ten different books, none of which was treated separately because of the small sample of students using each.

---

grammed Book (Prentice-Hall, Inc., 1966) or R. E. Attiyeh, K. G. Lumsden, and G. L. Bach, *Macroeconomics: A Programmed Book* (Prentice-Hall, Inc., 1967). Approximately two-fifths of the students in Group I used the Bingham text; one-third of the students in Group II used the Bingham text. Others used one of the Attiyeh, Lumsden, and Bach texts. All programmed learning and test materials were provided free to participating schools. Each school was allowed to choose the programmed text preferred, subject only to the constraint that each book be used by a typical cross-section of schools.

[3] For permission to use these tests we are grateful to the Psychological Corporation (304 East 45th Street, New York, New York, 10017) and to Professor Rendigs Fels, chairman of the TUCE committee, whose paper at this session provides more information on the tests. Other members were G. L. Bach, William G. Bowen, Paul L. Dressel (Executive Director), R. A. Gordon, Bernard F. Haley, Paul A. Samuelson, John M. Stalnaker (Consultant), and George J. Stigler.

[4] Participating colleges and universities, some of which participated in both micro and macro portions of the study, were the following: "High-prestige" schools: Amherst, Dartmouth, Harvard, Haverford, Oberlin, Swarthmore, and Yale; liberal arts colleges: Allegheny, Bowdoin, Denison, Lafayette, Union, and Wesleyan; large universities: Case-Western Reserve, City College of New York, Illinois, Indiana, Iowa State, Johns Hopkins, Minnesota, North Carolina, Oklahoma, Oklahoma State, Pennsylvania, Rochester, University of California, San Diego, Utah, and West Virginia; state (teachers) colleges: Arizona Western, California State, Fullerton, Georgia State, New Mexico State and San Diego State; junior colleges: Allegany, Bakersfield, Chabot, Cochise, Everett, Hill, Phoenix, Riverside City, Robert Morris and Triton; norm group: Butler, Geneva, Montana, State University of New York at Albany and Virginia Polytechnic Institute.

Results

The main findings of this study are reported in Table 2 which presents the statistics for a regression involving all of the variables discussed above. The dependent variable—the number of correct answers on the TUCE—had a mean of 17.90 and a maximum value of 30.[5]

All of the variables representing student characteristics were statistically significant[6] and quantitatively important. Each year of educational attainment added .39 points to the predicted grade, while being a girl cost .75 points. Not unexpectedly, student ability was the most important single determinant of student performance. Each 100 points on the composite college entrance exam (SAT scale) was worth 1.03 correct answers on the TUCE exam.

The results for school characteristics were somewhat more surprising. First, the average freshman entrance exam score was highly significant and nearly half as important to an individual's performance as his own entrance exam score. Each 100 points of the average SAT score of the freshman class added .45 correct answers to the student's score, even after all other individual student qualities are taken into account. It is impossible to tell whether this gain was due to associating with bright students or attending a high quality school that attracts bright students.

Second, school size had a positive, statistically significant coefficient; each 10,000 in enrollment added 0.2 points to the predicted score. Since school size in our sample ranged from 500 to 45,000 students, going to the largest university rather than the smallest school added approximately .9 of a question to student performance. We have no satisfactory explanation. It may simply be that larger schools have better courses and/or better instructors as measured by the TUCE. Or, perhaps, students who attend larger schools put their talent to better use.

Finally, of the school type variables, prestige schools and state colleges had large and statistically significant coefficients. Other things the same, compared to "other" schools, attendance at a "high prestige" school cost .78 points while being at a state college added 1.30 points. Whether these coefficients reflect differences in teaching effort or some other pertinent factor is difficult to say. In evaluating how a student would perform at schools of different types, however, it should be borne in mind that the quality (as measured by the average freshman SAT score) and size of these schools must also be taken into account. Leaving these factors free to vary, the influence on test score of attending a school of each type, as compared to "other" schools, is as follows: prestige schools, +.57; large state universities, +.71; liberal arts colleges, +1.08; state colleges, +.83; and junior colleges, −.90. These numbers still suggest that a student would perform better on the TUCE if he had gone to a less prestigious liberal arts college, a state college or a large university than if he had gone to a prestige college or university.[7]

Variables 11–13 show the difference in

---

[5] Actually, the TUCE consists of 33 questions. On the macro forms, however, we deleted from each form 6 questions dealing with microeconomics, leaving 27 questions. Two tests, therefore, had 33 questions each and two had 27 each. Thus, for all four forms, the unweighted average maximum score was 30.

[6] A confidence level of 95 percent will be assumed throughout the following discussion.

[7] A similar clasisfication was used in the study by G. L. Bach and Phillip Saunders entitled "Lasting Effects of Economics Courses at Different Types of Institutions," *A.E.R.*, June, 1966. In this study the significant finding was that the largest lasting effects for elementary economics courses were traceable to liberal arts colleges and the least lasting effects to large universities. Those results measure performance eight years after high school economics teachers took the elementary course.

TABLE 1

REGRESSION RESULTS
Dependent variable: test score (mean=17.90)
Coefficient of determination=.45
Standard error of estimate=3.51

Independent Variable	Mean	Regression Coefficient	"t" Statistic
Intercept		.28	
1. Educational level (year)	1.94	.39	5.34
2. Sex	1.16	−.75	−5.01
3. Entrance exam (SAT, hundreds)	11.08	1.03	28.37
4. Average school entrance exam (SAT, hundreds)	11.22	.45	5.95
5. School size (thousands)	10.62	.02	2.33
6. "Prestige" schools	.16	−.78	−2.40
7. Large state universities	.40	−.03	−.11
8. Liberal arts colleges	.13	.28	.92
9. State colleges	.08	1.30	4.54
10. Junior colleges	.15	.00	.00
11. Microeconomics test A	.24	.55	3.13
12. Microeconomics test B	.27	2.47	14.45
13. Macroeconomics test A	.27	−.31	−1.94
14. Conventional text A	.27	.86	4.68
15. Conventional text B	.09	.42	1.73
16. Conventional text C	.10	.75	2.89
17. Years teaching experience	4.67	.01	.26
18. (Years of teaching experience)^2	68.96	−.00	−.49
19. 1+class size	.02	.30	.06
20. P. L. book A+conventional	.26	.55	3.23
21. P. L. book B+conventional	.13	−.37	−1.72
22. P. L. book A	.19	−.02	−.07
23. P. L. book B	.12	−.46	−2.02

performance on the four test forms used. The standard of comparison is macroeconomics test B. Thus macroeconomics test A is .31 of a question more difficult. Since the microeconomics tests, however, each had 6 questions more and since students on average answered more than 50 percent correct (i.e., 17.9 out of 30), we would expect, if the tests were homogenous, that the coefficient on each of the micro tests would be greater than 3.0 (i.e., the average student would be correct on just over 3 of the extra 6 questions). Thus microeconomics test A is considerably more difficult than macro B which was the easiest of all the tests. It should be recalled that preliminary forms of the tests were used.

A surprising result is that teaching experience and class size have no observed effect on student performance.[8] What textbook is used, however, does appear to be a matter of some importance. Both texts A and C had positive significant coefficients. Other things equal, students who used these texts scored .86 or .75 points higher than students who used the average book in the "all other" category. There is no

---

[8] It should be noted that the result holds even though we tested for nonlinear relationships. We recognize, however, that experience is not necessarily a good proxy for quality of teaching.

TABLE 2

PROGRAMMED LEARNING COMPARED TO CONVENTIONAL INSTRUCTION

Alternative to Conventional Instruction	Difference in Predicted Score (P.L.-average conventional)	"*t*" Statistic*
P.L., only, book A...........	−0.54	−1.78
P.L., book A and conventional	0.55	3.23
P.L., only, book B...........	−1.02	−4.18
P.L., book B and conventional	−0.37	−1.72

* Calculated from Table 2 and the underlying variance-covariance matrix of the regression coefficients.

significant difference, however, between either books A or C and book B.

In evaluating the effectiveness of programmed learning taken by itself, the results are somewhat difficult to read from Table 1. The coefficients shown in Table 1, −.02 for P.L. book A and −.46 for P.L. book B, reflect the difference between programmed learning and a conventional course, assuming that variables 14–19 have a value of zero. In the case of textbooks, for example, this means that students in Group III are assumed to be using a text other than books A, B, or C. To make a comparison with the typical conventional course, it is necessary to give appropriate weight to the coefficients for variables 14–19 in Table 1. This comparison is shown in Table 2. What these figures indicate is that, other things the same, students using programmed learning did half a question worse, in the case of P.L. book A, and one question worse in the case of P.L. book B, than did students in the average conventional course. In the case of P.L. book A, however, this difference is not statistically significant.

As for the effectiveness of programmed learning when it is used as a supplement, P.L. book A added significantly, about half a question, to student performance, whereas P.L. book B did not. These results are shown in both Tables 1 and 2.

The questions on the TUCE were classified into the following categories: recognition and understanding; simple application; and complex application. Questions in the first category required knowledge of relatively more historical and institutional material, while questions in the latter two categories usually stated all the necessary factual material and tested the ability to use economic analysis. Separate regressions were run to explain student performance on each type of question. The statistics for these regressions are shown in Table 3 and their implications for the comparison between programmed learning and conventional instruction are shown in Table 4. The most striking result is that, for P.L. book A, most of the overall difference between student performance in Groups I and III was concentrated in the recognition and understanding category. Since the programmed learning texts were designed primarily to teach basic tools and not all basic introductory and institutional material covered by the typical course or the TUCE, the differences in this category are not surprising. For those questions requiring an ability to apply economic theory, however, students who used P.L. book A only for 12 hours, on average, were an even match for conventionally taught students.

In addition to the objective information already discussed we solicited student reactions to programmed learning. The re-

TABLE 3

Regression Results by Type of Question

Independent Variables	Recognition and Understanding		Simple Application		Complex Application	
	Regression Coefficient	"t" Statistic	Regression Coefficient	"t" Statistic	Regression Coefficient	"t" Statistic
Intercept	1.35		−.05		.66	
1. Educational level (year)	.12	3.91	.12	3.67	.19	5.22
2. Sex	−.34	−5.44	−.21	−3.02	−.27	−3.50
3. Entrance exam (SAT, hundreds)	.28	18.20	.37	22.12	.38	20.15
4. School entr. exam. (SAT, hundreds)	.12	6.08	.08	3.80	.11	4.52
5. School size (thousands)	.01	4.63	.01	2.29	.00	1.34
6. Microeconomics test A	.04	.62	1.33	17.15	−.75	−8.51
7. Microeconomics test B	−.07	−.96	2.57	34.41	−.07	−.85
8. Macroeconomics test A	.83	12.37	−.19	−2.64	−.96	−11.46
9. Conventional text A	.26	3.47	.14	1.71	.27	2.92
10. Conventional text B	.28	2.88	.10	.97	−.26	−2.18
11. Conventional text C	.31	3.12	−.05	−.43	.07	.59
12. P.L. book A+conventional	.19	2.65	.14	1.76	.21	2.41
13. P.L. book B+conventional	−.09	−1.03	−.14	−1.54	−.10	−.94
14. P.L. book A	−.24	−2.62	−.05	−.49	.10	.92
15. P.L. book B	−.15	−1.76	−.24	−2.53	−.27	−2.46
	$R^2 = .31$ $S = 1.47$		$R^2 = .45$ $S = 1.61$		$R^2 = .31$ $S = 1.83$	

sponses to the questions concerning effectiveness and interest are summarized in Table 5. Several aspects of these distributions stand out. First, the mean response was favorable, but there was a wide dispersion of opinions. Second, students considered programmed learning to be somewhat more effective than interesting. On average, the students gave the books grades of "good-minus" for effectiveness and "average-plus" for interest. Finally, although the objective evidence revealed a

TABLE 4

Programmed Learning Compared to Conventional Instruction, by Type of Question

Difference in predicted score* (P.L.—average conventional)

Alternative to Conventional Instruction	Recognition and Understanding "t"		Statistic Application "t"		Complex Application "t"	
	Difference	Statistic	Difference	Statistic	Difference	Statistic
P.L. only, book A	−.42	−4.29	−.11	−1.03	.02	.17
P.L. A and conventional	.19	2.65	.14	1.76	.21	2.41
P.L. only, book B	−.34	−3.58	−.32	−3.08	−.36	−3.05
P.L. B and conventional	−.09	−1.03	−.14	−1.54	−.10	−1.13

* Calculated from Table 5. The "t" statistics are calculated using information from the variance-covariance matrix of the regression coefficients for the regression reported in Table 3.

TABLE 5

STUDENT ATTITUDES TOWARD PROGRAMMED LEARNING*

As a way of learning economic theory what do you think of the programmed text?

	Group I	Group II
Very good	20%	18%
Good	42	43
Average	23	25
Poor	12	10
Very poor	3	4

How interesting was the programmed text?

	Group I	Group II
Very interesting	6%	3%
Interesting	41	32
Average	34	42
Uninteresting	12	15
Boring	6	7

* Figures may not add to 100 because of rounding or failure to receive answers.

significant difference between the two books, this was not reflected in the student's opinions about them.

Conclusion

We feel that these results have important implications for the organization and teaching of the introductory course. Within the profession many believe that the introductory course should prepare a student to think intelligently about major economic problems in modern society and that this goal can best be accomplished by teaching a few basic principles and applying them to a number of important problems. We are in agreement with this view. This study has shown that by using programmed learning materials the basic micro- and macroeconomic theory can be taught in a relatively short period of time. Therefore, more time can be devoted to teaching students how to apply the theory to social problems, both by going more deeply into the more important problems and by actually covering those topics scheduled for the end of the term that often fall victim to the school calendar. The use of these materials can have other advantages: First, the student can gain a good overview of the entire course at the very beginning which helps him to put topics covered in the remainder of the course in meaningful perspective. Second, because a course taught in this manner emphasizes the usefulness of economic theory in a problem solving context, it promises a positive impact on the most important single factor in the learning process; namely, student attitude toward the subject.

PERFORMANCE ON THE NEW TEST OF UNDERSTANDING IN COLLEGE ECONOMICS

By Arthur L. Welsh, *Joint Council on Economic Education*
and
Rendigs Fels, *Vanderbilt University*

Nearly three years ago the Committee on Economic Education of the American Economic Association recommended that the Joint Council on Economic Education commission the preparation of a test for use in the college level elementary economics course. The Joint Council appointed a Committee to prepare the test and that task has now been completed.[1] The Test of Understanding in College Economics, as the newly completed tests are collectively called, consists of a pair of tests for each semester of the typical elementary course. Two versions (designated A and B) for each semester make possible before-and-after testing of groups of students on a comparable basis. Each of the four forms was designed for administration in one class hour.

More than four hundred multiple-choice questions were tried out in college classrooms between January, 1967, and January, 1968. During tryout each question was answered by several hundred students of elementary economics in schools representing a cross-section of American higher education. Questions were included in the test only if the tryout results showed that they discriminated reasonably well between good and poor students. Questions were rejected for which the correlation between the number of right answers and the quality of students was negative or low. Rejected also were questions that proved unduly difficult or unduly easy.

The primary purposes of TUCE are (1) to serve as a research instrument for controlled experiments and (2) to provide an evaluation instrument for the elementary economics course in college. In addition to use of the test by Bach, Lumsden, and Attiyeh in the experiment reported on at this session, several other investigators are using the TUCE for experimental purposes.

In norming the TUCE a great deal of data was collected on how well students performed on the test, and it is some of the results of student performance we propose to present here. In collecting norming data during the academic year 1967–68, Part I, dealing mainly with the macro aspects of the elementary course, was given to a group of students at the beginning (pre-test) and at the end (post-test) of the course. Part II, mainly microeconomics, was given only at the end of the course.

The results of the test for which we made rough comparisons between "good" and "poor" students only are interesting. Good students are defined as the 27 percent of the total group tested who scored highest. Poor students are the 27 percent of all students tested who scored lowest. Perhaps the most comforting of our findings is that students know more economics after a semester of economics than before. Whether or not everyone will be satisfied with the amount of economic knowledge acquired is, of course, quite another matter. Students scoring highest on the pre-test—good students—managed to answer, on the average, 18.5 questions out of the total of 33 correctly. On the post-test, that is, at the end of the semester, students scoring highest were able to answer slightly more than 25 questions correctly. At the other end of the scale the lowest 27 percent of all students answered about 9 questions correctly on the pre-test—a figure little more than one would expect if the students only guessed at the answers. At the end of the semester the students who scored most poorly on the post-test managed to answer about 12.5 questions correctly out of a total of 33.

---

[1] The Committee appointed by the Joint Council on Economic Education to prepare the TUCE included: Rendigs Fels, Chairman, Paul L. Dressel, Executive Director, G. L. Bach, William G. Bowen, George J. Stigler, Bernard F. Haley, R. A. Gordon, Paul A. Samuelson, and John M. Stalnaker, Consultant.

Comparisons of good and poor students on a before-and-after basis cannot be made with any certainty on the strength of data we have processed thus far. Those who scored highest on the pre-test are not necessarily the same ones who scored highest on the post-test. The same holds true for the lowest group; low scorers on the pre-test are not necessarily low scorers on the post-test. Despite this obvious shortcoming in our data, the following points are worth noting. (1) The fact that 27 percent of our students can answer little more than one-third of the questions on the test after an entire semester of economics is disturbing to say the least. It might well be true, as frequently hypothesized, that we try to do far too much in the elementary course and consequently accomplish very little. Of course some might argue that the fault lies with the test instrument and not with the course itself. Possibly. But more on this later. (2) The top 27 percent of all students on the pre-test managed to answer more than half the test questions correctly. Moreover, their average score at the beginning of the course was almost as high as the average for all students at the end of the first semester. Since the top 27 percent on the pre-test probably had scores above average on the post-test, there is a good chance that their scores before instruction exceed those of all the other students after instruction. This might indicate that some students, at least, are entering college with a fair background in the subject. If this is indeed the case—and it seems worth investigating—then a pre-test of all new students could be useful in planning the college course.

The overall mean scores of all participating schools on Part I were 13.55 and 19.16 on the pre-test and post-test, respectively. On the pre-test, mean scores for individual colleges ranged from 10.76 to 18.00. It might be noted here that the score of 18.00 is atypical. If this score is eliminated, 22 of the 23 schools in the test sample had mean scores ranging from 10.76 to 14.88. When we look at the range of post-test scores, 22 of the 23 means fell in the range of 15.78 to 22.83. The school on the post-test which had the lowest mean score of 12.67 was the same institution scoring lowest on the pre-test.

On the average, the improvement factor in all institutions for Part I was about 41 percent. Improvement ranged from almost 70 percent in one institution to approximately 18 percent in the lowest. Pre-test data for Part II are unavailable and no comparisons could be made with post-test results. However, the mean score on the post-test of Part II was 18.63.

Institutions of higher education vary greatly in terms of student body, faculty, and educational emphasis and one is therefore cautioned against placing too much emphasis on interinstitutional comparisons. Nevertheless, for those interested in comparisons, average scores give a rough approximation of how well students in one institution are performing vis-à-vis all others.

One must be cautious about drawing inferences from the data for individual questions. Although the norming data have been derived from a large enough sample so that the orders of magnitude are probably about right, a variety of considerations affect the outcome. The same question will yield different results from the same students when asked in different contexts. Not only do results differ when the questions in the examination are different but also when the other questions are the same but the order is different. An unknown proportion of the right answers comes from guessing. Since answering a multiple-choice, four-option question is not a random process, we cannot simply deduct 25 percent from the proportion answering the question correctly to allow for guessing. The guessers may all be attracted by a particular distractor. The percentage of correct responses depends not only on the difficulty but also on the quality of the question. Any latent ambiguity in the question will lower the score. In what follows, we shall look at certain questions and try to limit ourselves to inferences that seem reasonable. However, our inferences must be regarded as no more than suggestive.

The Test Committee agreed early to limit the number of questions which involved only rote knowledge or recall. Instead the Committee felt that the majority of the questions should test students' ability to apply basic economic principles and concepts to realistic situations. Hence the Committee decided upon three classifications of questions: (1) recognition and understanding—questions which in principle could be answered by reference to a textbook, although even here the better questions involve the restatement or recognition of an idea in somewhat different terms than it was originally learned; (2) simple application—questions in which the student is called upon to use a principle or concept when it is clearly specified or implied; and (3) complex application—questions requiring students to select and use a principle or concept when its use was not specified or requiring two or more principles or concepts related in some manner not previously presented to students.

Two questions on Part I of the TUCE (I-A-32 and I-B-33) test knowledge of a simple but basic concept: the meaning of a demand schedule or curve. On both questions, the percent of all students choosing the right response is very low: 42 percent and 25 percent. Although both questions are classified as simple applications, they are such straightforward questions that they might have been called "recognition and understanding." The one on which scores are lower is negatively worded ("Which of the following could NOT have an effect on the demand curve. . . ."). Since negatively worded questions tend to be confusing, the higher of the two scores is the better index of student knowledge. It indicates that at the end of a semester of economics only two-fifths of college students understand what a demand schedule is.[2] The low scores may mean that some schools in the sample from which the norming data came make no effort at all to teach supply-and-demand analysis in the first semester.[3] But since some supply-and-demand analysis is a normal part of the work of the first semester, the data suggest that we are not succeeding in teaching the fundamentals of the subject. Further investigation is needed.

---

[2] A third question on this concept (I-B-6) at first sight seems to give contradictory evidence, since 75 percent get it right at the end of the semester. But an even higher proportion—81 percent—get it right at the beginning of the semester, suggesting that the question is so worded that students can get it right without understanding the concept. From a psychometric point of view, I-B-6 is an inefficient question and should be replaced when a new edition of the TUCE is published. Unfortunately, when the test was constructed, we had no data to show that students in large numbers could get it right without knowledge of the concept being tested. The question does no harm in the test; the objection is that it contributes little toward measuring student achievement.

[3] Since the TUCE was usually administered before the final examination, the scores may have suffered because the students had not yet done their final intensive reviewing. On the other hand, forgetting sets in at a rapid rate after the course is over. The score on I-A-32 is suggestive rather than conclusive, but it gives a rough indication of how much (or rather how little) students are learning.

Each of the forms of Part I has a question on a basic aspect of the allocation of resources. One of the questions (I-A-24) is on the nature of the economizing problem; the other (I-B-32), on the price mechanism in its simplest workings. Although both are classified as simple applications, both appear to be so elementary that they ought to be easy if we professors are doing our jobs right. Yet the scores are low: 47 percent and 31 percent of all the students in the national sample chose the correct responses. In Part II there is another basic question on the price mechanism (II-B-5). It is even simpler and is classified as "recognition and understanding." The score at the end of the second semester is quite respectable: 77 percent. Yet it indicates that even after a full year of college economics, nearly a quarter of the students do not understand the ABC's of a market economy. Study of the distribution of the wrong answers shows that most of the confusion is over the role of profits. Although more data are needed to verify these results, the data strongly suggest that we are neglecting an elementary part of our subject.

The percentage of students getting question I-A-3 right is higher at the beginning of the first semester (85 percent) than at the end (80 percent)—a disconcerting result. The question is a simple one of the recognition-and-understanding type, testing whether students understand the meaning of "real wages." Although the retrogression may represent a sampling error,[4] and although the percent getting it right at the beginning of the semester is so high as to leave little room for improvement, the data suggest that instructors are slighting an important concept, regarding it as too simple to pay any attention to.

With respect to technical economic theory, the TUCE provides mixed evidence as to how much we are achieving. Question B-2 and A-7 on Part II of the TUCE show that when students are confronted with a simple numerical example of supply and demand, most of them can pick out the equilibrium price and quantity (the percents of correct responses are 88 percent and 85 percent) and II-A-8 indicates that about three-quarters of them realize that setting a price above equilibrium results in producers not being able to sell as much as they would like. So far, so good. On the other hand, only about two-fifths of the national sample knew that if marginal cost is less than marginal revenue, the firm can increase profits by expanding output irrespective of whether competition is perfect or imperfect (II-B-24). Once again, the result suggests that we may be misallocating our efforts, spending too much time drawing curves and too little on fundamentals.

Question II-B-21 points in the same direction. It is a variant of an old chestnut. A synthetic "quotation" argues that if there is excess demand, the resulting rise in price will lower the demand curve and raise the supply curve. The norming data indicate that three-fifths of the students tested recognize that the quotation confuses shifts of the curves with shifts along them, but more than one-fifth think the quotation is correct.

In some cases, the fact that a question under consideration for the TUCE had to be discarded is revealing. One such question was derived from the following true-false question used by Simon Whitney: $1.00 deposited in a banking system keeping 20 percent reserves will add \$5.00 to loans and investments.[5] For three reasons the statement is false. Yet, Whitney found that more students thought it true *after* taking an economics course than before. We rewrote Whitney's question several times in multiple-choice form. The results were never satisfactory. The version used for the tryout for the TUCE read:

---

[4] On any one question a 4 percent difference in right answers is significant at the 5 percent level, a 5 percent difference at the 1 percent level.

> One dollar deposited in a banking system that always keeps 10 percent reserves and suffers no loss of currency will *add* how much to the system's loans and investments?
> 1. \$0.90
> 2. \$5.00
> 3. \$9.00
> 4. \$10.00

On the tryout, the number getting it right was only 25 percent, the correlation between a right answer to this question and scores on the tryout test as a whole was marginal (0.29), and the correlation was especially unfortunate for the \$10.00 option (one-third of both the top 27 percent and the middle 46 percent chose that response, indicating that it was not discriminating effectively between good and average students). In no one of these respects was the question too bad to use, but all three of them together induced us to discard it. We might have had a usable question if we had used \$1.00 for a response rather than \$10.00. Two inferences seem warranted: (1) questions of this type are too tricky to be used with elementary students and (2) instead of getting elementary students to understand the model of money creation, we may be instilling in them little more than a reflex action: when \$1.00 goes in, \$5.00 will come out if the reserve ratio is 20 percent, or \$10.00 of it is 10 percent. Economics by the numbers, as it were.

Three questions used in the TUCE are closely similar to questions which appear on the Test of Economic Understanding[6] designed for high school use. In one case (TUCE I-A-27 and TEU-B-36), where the wording was identical except for an insignificant difference in the stem, the high school students used for norming the TEU scored almost the same (43 percent right) as the college students used for norming the TUCE (45 percent). The question calls for naming a policy that would not increase growth when the economy is at full employment, the right answer being a policy that would increase consumption spending. The negative wording of the question accounts in part for the low scores. Since college students scored lower on this question before taking college economics than high school students before taking high school economics (31 percent and 36 percent, respectively), college teachers can comfort themselves that their students appear to have improved more (14 percent compared to 7 percent). The most popular wrong response is encouraging savings and investment (50 percent before college economics, 39 percent after). Perhaps the effect of teaching Keynesian economics, with its emphasis on how saving may depress output (and therefore the current growth rate), hinders greater improvement in answering the question. Nevertheless, after all the rationalizations we have been able to think of, the failure of college students to do better is disturbing.

A question on the effect of a fall in total demand—the right answer being a slower growth rate, the wrong answers inflation, more employment, and a government surplus—appears in the TUCE (I-A-4) and TEU (B-27) with different wording. The college students, according to the norming data, do better than the high school students by 10 percentage points (66 percent and 56 percent). The margin is nothing to be proud about for such a simple question —one-third of the college students missed it. Furthermore, nearly half (46 percent) got it right before taking economics. Thus, the data suggest that more than half of the students who could not answer the question before taking economics still could not answer it after studying economics for a semester.

---

[5] Simon N. Whitney, "Measuring the Success of the Elementary Course," *A.E.R.*, May, 1965, p. 567.

[6] Joint Council on Economic Education; Science Research Associates, Chicago, 1964.

On the question of what part of total spending fluctuates most during business cycles, appearing with different wording in the TUCE (I-B-25) and the TEU (A-29), the college students did worse than the high school students, scoring lower both before (13 percent and 20 percent, respectively) and after (28 percent and 39 percent) studying economics. The lower scores for college students presumably result from the different wording and the different context. That the percentage was so low at the end of the course—hardly more than would be expected from sheer guessing—must mean that college teachers are no longer teaching about business cycles.

Whether we look at overall test results or results on individual questions, for many of our students economics evidently is difficult, and we may be expecting too much from them. Principles and concepts deemed too simple to spend much time on are not at all obvious to students even at the end of the semester.

The TUCE itself is far from perfect and no one who has been closely associated with its developments would dispute this point. Some may find fault with an attempt to draw up specifications for a "typical" course where none in fact may exist; some may object to the use of multiple-choice questions which presumably "preclude discernment of gradation and sophistication in the answers," and some may find the questions too difficult, too trivial, or both. To these objections we admit to some degree of truth.

But the elementary course engages a considerable amount of the resources in most economics departments. It is also an area in which economists who deplore casual impressions in research have relied on little more than casual impressions in evaluating teaching. For research purposes or for diagnosing the effectiveness of the current teaching, the TUCE represents the only easily available instrument for which there are data representing a good cross-section of American college students. Our modest hope is that the TUCE will encourage more research in a badly neglected area.

DEEP: STRENGTHENING ECONOMICS IN THE SCHOOLS

By JOHN E. MAHER
Joint Council on Economic Education and *State University of New York, Binghamton*

This paper is a report on the impact of the Joint Council on Economic Education's five-year effort to advance the teaching of economic analysis in major school systems throughout the nation. The program, known as DEEP (Developmental Economic Education Program) reached into thirty systems enrolling 4½ million pupils. I believe economists have a great stake in the endeavor: (1) because, as teachers, they may find that the instruction of their own students in college requires modification in view of the increasing economic knowledge of students who have already encountered their discipline; (2) because, as educators, they may find helpful inferences for handling their own educational tasks; and (3) because, as consultants, they are going to be asked with increasing frequency to help school systems with the job of strengthening the economic dimension of social studies.

The thrust of DEEP is not principally to add separate courses in economics to school curricula but to improve the economic analysis that is integral to the teaching of social studies. Because teachers must attempt to handle the economic dimension of free silver, the Great Depression, and like topics, there can be little room for debate as to whether they ought to be prepared to handle them.

General Objectives of DEEP and Measures of Their Attainment

Recent studies have suggested the proximate reasons for the lack of good instruction in economic analysis in the schools: inadequately prepared teachers, a dearth of appropriate teaching materials, and, at a more derivative level, lack of consultative help from the economics profession.[1]

DEEP seeks to meet these deficiencies by establishing partnerships among school systems, universities, state councils for economic education, and the Joint Council. The purpose of these partnerships is to create courses of study in economics for teachers—a kind of on-the-job (in-service) training—and to develop course guides and pupil readings stressing economic analysis.

From this brief statement of the objectives of DEEP we can formulate a few questions about the degree of their attainment: (1) How successful was the program in getting the essential teaching and consultative input from professional economists? (2) How many teachers received instruction in economics and with what effect on their understanding of the discipline? (3) What was the output of curriculum materials developed for teachers and their students? (4) What effects upon student understanding of economics have re-

---

[1] Only about half of the nation's 60,000 social studies teachers have ever undertaken formal study of economics. See G. G. Dawson, *Nation-Wide Survey on Economic Education of Teacher-Trainees* (Joint Council on Economic Education, New York, 1967). Of the 64 percent of the colleges requiring economics of some prospective teachers, only six or fewer semester hours fulfill the requirement. G. L. Bach and P. Saunders state that there is little lasting effect from study in fewer than five courses, so the present level of requirements appears to be insubstantial [1]. Surveys in DEEP show that about two-thirds of elementary teachers were untouched and one-third of the junior-senior high school teachers. A 1968 survey by the Joint Council on state requirements for the certification reveals that twenty-two states require no economics of high school teachers of economics; thirty-six states require none of high school social studies teachers; thirty-eight states require none of junior high school social studies teachers; forty-seven states require none of elementary.

TABLE 1

DEEP Teacher Education, 1964–65 Through 1966–67

Number of Teachers Enrolled						Total Enrollment	Total Employment
By Grade Taught			By Academic Year				
K-6	7–12	Others	1966–67	1965–66	1964–65		
4,226	2,788	304	7,318	4,595	1,779	13,692	174,922

sulted from the program? In this section, general answers are suggested to the first three questions with the fourth question deferred to the next section.

The Participation of Economists. DEEP was designed by the Joint Council in consultation with an Advisory Committee of the A.E.A.'s Committee on Economic Education with frequent reporting to and feedback from the Advisory Committee. The effect of this consultation has been large. I believe the major influence of these representatives of the profession has resulted in better optimizing: the more explicit definition of performance criteria and the systematic attempt to measure the extent of the attainment of objectives and the degree of efficiency realized.

Thanks in part to the concern of the A.E.A., about seventy-five economists from forty colleges and universities have taught and consulted with the thirty school systems for periods of up to four years. In three systems, economists were in full-time residence (i.e., on the payroll) and in other systems they devoted the academic equivalent of about quarter-time. This participation is a remarkable step forward in strengthening the teaching of economics in the schools.

The Education of Teachers. In the first three years, nearly 14,000 teachers were instructed in economics. It must be added that while this instruction sometimes amounted to a two-semester university course, in more instances it consisted of a summer workshop or three-day institute.

Table 1 shows the distribution of teachers instructed in the thirty school systems during the first three years of DEEP.

The weighted mean percentage of teachers instructed is 7.8 percent. The unweighted mean which more nearly represents the typical school system is 14.9 percent.[2] If only those ten systems which had completed three years' work are tallied, the weighted mean becomes 12 percent and the unweighted mean, 24.7 percent. Assuming that such offerings continue so that the hiring of new teachers and the loss of those more experienced does not dilute the level of preparation in economics, the median DEEP school system after three years will have raised the proportion of teachers who have been exposed to the subject from the national average of 50 percent to an average of about 75 percent.

With what results for their knowledge did these several thousand teachers study economics? For courses of a semester's duration, a sizable sample of before-and-after testing with the Test of Economic Understanding provides ground for qualified satisfaction. (For a discussion of the

---

[2] Most of the difference between means is accounted for by New York City and the state of Wisconsin which employ nearly half of the teachers.

TABLE 2

MEASURED CHANGES IN ECONOMIC KNOWLEDGE OF TEACHERS ENROLLED IN ONE-SEMESTER COURSES, 1964–67

Enrollment	School System	Mean Pre-test Score, Form A or B	Standard Error of the Mean	Mean Post-test Score, Form A or B	Standard Error of the Mean
60	R	24.8 (A)	n.a.	33.5 (B)	n.a.
11	S	35.3 (B)	2.02	38.1 (B)	1.84
40	S	35.1 (B)	1.09	39.5 (B)	1.09
26	S	37.0 (B)	1.61	42.2 (B)	1.06
68	S	34.8 (B)	.72	42.0 (B)	.65
87	T	30.4 (A)	n.a.	37.3 (B)	n.a.
42	U	29.5 (A)	n.a.	35.0 (B)	n.a.
42	V	36.1 (A)	.99	43.0 (B)	.79
				Weighted Average 38.5	
376					

test, see [1].) Table 2 shows test results for 376 teachers in five school systems. Except for System S, $N = 11$, statistically significant improvements in economic understanding occurred at the .001 level as measured by the test (assuming the missing errors of the mean are no larger than the largest error reported). Statistical significance is not, however, a good criterion because small changes in knowledge may so easily be statistically significant.

A more meaningful assessment of the state of teachers' knowledge can be derived by comparing their mean weighted post-test score of 38.5 with the means reported in other studies. Thus Bach and Saunders [1] report a mean of 41.2 for teachers who regularly watched the TV course, *The American Economy,* and a mean of 40.7 for college sophomores at Carnegie-Mellon and the University of Nebraska; Moyers and Paden [4] report a post-test mean of 42.0 for students at the University of Illinois; and McConnell [3] reports means between 38.7 and 39.6 in later work at Nebraska. While the mean for DEEP teachers of 38.5 is significantly lower than most of these other scores, the differences are not large. It is fair to conclude that according to the TEU these teachers show nearly as much understanding of economics at the end of their course of study as the teachers and college students reported by others.

The Development of Materials. It seems useful here only to allude to hundreds of pieces of instructional material developed by DEEP school systems and their university partners: readings for students, teachers guides, TV tapes, and other things. These vary enormously in both economic content and in educational quality. Twenty of the written items, collectively spanning the grades from one to twelve, have successfully passed the Joint Council's criteria for national dissemination and are now in widespread use.[3] As with our assessment of the education of teachers, so also with curriculum materials we must ask: Have these materials had palpable effect upon student knowledge? The next section is devoted wholly to this question.

Impact on Student Knowledge

Testing of student knowledge with the TEU has followed two avenues: (1) gen-

---

[3] A brief annotated listing appears on the Checklist, J.C.E.E., 1968. For more detail on the process of developing materials and the criteria used in their evaluation, see *DEEP'67* (J.C.E.E., New York, June, 1967).

TABLE 3

LARGE-SCALE TESTING WITH TEU OF SENIOR HIGH SCHOOL STUDENTS

	System	Time of Year	Grade	Population	Individual		Standard Error
					Mean	Form	
First year of participation	R	Fall '64	10	8502	18.4	A	.06
	S	Fall '64	10	3442	18.2	A	n.a.
	T	Fall '64	10	1739	13.55	B (scaled)	.15
	R	Spring '64	10	411	18.4	B	.27
	U	Spring '66	10	1081	12.2	A	n.a.
	R	Fall '64	11	7833	22.3	A	.07
	S	Fall '64	11	2736	18.46	A	n.a.
	R	Spring '65	11	383	19.4	B	.37
	U	Spring '66	11	1507	12.7	A	n.a.
	R	Fall '64	12	6470	23.1	A	.09
	S	Fall '64	12	2692	20.14	A	n.a.
	V	Fall '65	12	1857	20.46	A	n.a.
	T	Fall '64	12	1339	23.06	B	n.a.
	R	Spring '65	12	4157	23.5	B	.12
	S	Spring '65	12	2783	22.70	B	n.a.
	V	Spring '65	12	1752	25.43	B	n.a.
	T	Spring '65	12	1339	26.31	B	n.a.
	U	Spring '66	12	1571	14.35	A	n.a.
Second year of participation	V	Fall '66	12	1835	19.89	A	n.a.
	W	Fall '66	12	1387	21.9	A	.18
	S	Spring '66	12	2754	22.12	B	.15
	V	Spring '67	12	1704	23.63	B	n.a.
Third year of participation	W	Spring '67	12	1211	23.9	B	.21
	R	Spring '67	12	653	25.52	B	n.a.
	S	Spring '67	12	2811	21.36	B	.14
	R	Spring '67	12	690	23.19	A	n.a.
			Total	64639			

erally large-scale testing of senior high school students to establish bench marks, so that later testing would permit an assessment of the gross impact of changes in curriculum and instruction; (2) smaller-scale testing of particular classrooms before and after their instruction in economics whether that instruction was part of a course in economics or in social studies.

Large-scale Testing. Data in Table 3, mostly of a cross-section nature, show results of testing over 60,000 students at 10th, 11th, and 12th grades. Taking what are nearly census results from a single system, say System R, we are now for the first time able to give a static profile of student knowledge. Mean scores reveal the following situation:

Grade	Mean Score	Intergrade Differences
10 (Fall)	18.4	3.9
11 (Fall)	22.3	.8
12 (Fall)	23.1	1.9
12 (Spring)	25.0 adjusted[4]	
		Sum 6.6

---

[4] Our studies indicate that for 12th graders, Form B

Interpreting these data dynamically, we may infer that over a three-year period students show a gain in knowledge of 6.6 points—about the same gain as might be expected from a semester's course in economics. It should be emphasized that this growth in knowledge results from a composite of such influences as student maturation, classroom experience, exposure to mass media—all the influences of nature and nurture. It must be added that the data are biased upward in the sense that at each grade level there remain to be tested only those presumably more talented or persistent students who have not earlier dropped out of school.

Notable also in System R is the fact that nearly two-thirds of the growth in knowledge occurs during grade 10 (3.9 out of 6.6 points). This contrasts markedly with System S where the converse occurs: two-thirds growth taking place during grade 12 (4.1 out of 6.0 points):

Grade	Mean Score	Intergrade Differences
10 (Fall)	18.2	.3
11 (Fall)	18.5	1.6
12 (Fall)	20.1	4.1
12 (Spring)	24.2 adjusted[5]	
		Sum 6.0

Test results from Minneapolis permit the longitudinal study of a constant population over grades 10, first semester, to grade 12, second semester, showing differences by sex, socioeconomic status, and enrollment in courses in economics.[6] The data shown in Table 4 suggest the superior performance of these "persisters" who started from a 10th-grade mean of 19.36 and rose to 28.06 (on the more difficult TEU test, Form B), a gain of 8.70 points. This gain contrasts with the profile "gain" cited above, of about 6 points. Moreover, even the lower socioeconomic group with no economics course offered at the school shows a gain of 6.82 points.

Another finding from the Minneapolis study is that, without exception, the higher the socioeconomic status the better the performance and the more the economic input, the better the performance. Thus, at 10th grade, before any study of the subject, every group in school A outscores every group in school B and every group in B outscores every group in C. By the end of grade 12, when all students in A have taken a course in economics, some in B have taken a course, and none in C, the performance gap has widened further.[7]

Additional study of these and related data, cross-sectioned and longitudinal, may permit better analyses of the impact of DEEP upon the economic knowledge of senior high school students.

Testing the Effect of Particular Materials and Instruction. The Pittsburgh public schools have shown the most tightly structured program for developing and evaluating teaching guides and student readings.[8] For grade 8 through 12, their effort includes the testing of an unrevised edition of student readings, the re-

---

of the test is 1.5 raw score points more difficult than Form A. Therefore, in the text above, the raw score shown in Table 3 of 23.5 has been raised to 25.0 to make it comparable to the other, Form A scores.

[5] See preceding note.

[6] Unpublished report to the Joint Council by Professor John D. Helmberger, formerly resident economist in the Minneapolis School System and now at the University of Minnesota. Data from three high schools cover all grade 10 students (1964) who reached grade 12 (1967).

[7] There are, of course, numbers of uncontrolled variables, among them quality of instruction and levels of ability and motivation. One reason for not considering the absence of control for these variables, a severe limitation, is that "social-economic status," from an educational viewpoint, may be defined as mostly a composite of these variables.

[8] Unpublished report to Joint Council by Phillip Saunders, of Carnegie-Mellon University.

TABLE 4

GROWTH IN ECONOMIC KNOWLEDGE FOR A CONSTANT POPULATION IN MINNEAPOLIS, 1964–67

Group Tested		Grade 10 Pre-test'64			Grade 12 Post-test'67			Change in	
		N	M	s.d.	N	M	s.d.	M	T-Test
School A	Boys	243	21.95	6.39	243	33.34	8.22	11.36	28.07
	Girls	229	20.62	5.47	229	31.28	7.57	10.66	27.44
	Both	472	21.30		472	32.34		11.04	
School B	Boys	142	17.91	5.74	142	25.78	7.57	7.87	15.79
	Girls	158	19.25	5.35	158	26.10	7.27	6.85	15.05
	Both	300	18.62		300	25.95		7.33	
School C	Boys	162	16.85	5.29	162	24.73	7.98	7.88	14.95
	Girls	208	17.98	6.08	208	23.99	8.00	6.01	15.85
	Both	370	17.49		370	24.31		6.82	
Total		1142	19.36		1142	28.06		8.70	

vision of the readings in light of test findings, and further testing of the material after revision.

Preliminary results of the two stages of testing are now available for grades 8 and 9. The following table summarizes the results by grade and stage: stage one refers to the tryout of unrevised material and stage two to the tryout of revised material.

Looking at grade 8, we see that test and control groups have nearly identical I. Q. scores, reading level scores, and economics test scores. After instruction, both groups show marked gains on the economics test but the experimental groups using specially prepared materials taught by specially trained teachers have a gain that is two and one-half times that of the control groups: 23.4 percent as against 9.4 percent.

More extensive testing of revised material at stage two permits evaluation of the effect of training since all groups used the same material but the control group lacked special training in its use. Here we find substantial advances in knowledge among all groups with the control group (without training) lagging significantly.

Similar results are obtained from grade 9. The gain in economics score at stage one for students using unrevised materials and taught by trained teachers is over two and one-half times as great as for students in control groups. At stage two with revised materials, test scores rise significantly. (Only results for the experimental groups are presently available.) While the gain is not much greater for the revised than for the unrevised material, it must be noted that the groups tested after using revised material began at a relatively much higher level of economic knowledge.

There is little doubt that, as Saunders suggests, students "learn some economics in history and social studies with untrained teachers and without special material, they learn more with untrained teachers with special material, they learn still more with trained teachers using special material. . . ."

The Wichita school system conducted an evaluation of the effectiveness of instruction in economics in a sample of 208 students, using as additional explanatory variables sex and school attended. The latter variable indicated whether or not

TABLE 5

EVALUATION OF MATERIALS IN PITTSBURGH

Group	Number of Students	Average I.Q.	Average Reading Level	Average Economics Score*		Change in Score
				Before	After	
Grade Eight (United States History) Stage One						
4 test groups......	178	102.7	7.6	29.8%	53.2%+	23.4%
5 control groups lacking materials and without special training.....	183	102.2	7.5	29.7	39.1 +	9.4
Stage Two						
24 Pittsburgh city teachers.........	1022	104.8	n.a.	38.0%	52.8%+	14.8%
24 parochial school teachers.........	1437	110.5	n.a.	41.6	56.4 +	14.8
5 county teachers..	392	113.4	n.a.	43.6	57.2 +	13.6
4 "untrained" but otherwise comparable teachers....	205	110.0	n.a.	40.5	51.6 +	11.1
Grade Nine (Social Studies) Stage One						
7 test groups......	300	100.7	8.4	39.2%	53.2%+	14.0%
7 control lacking materials and without special training.........	418	100.1	8.3	39.6	44.8 +	5.2
Stage Two						
8 experimental groups..........	548	107.0	n.a.	47.6%	62.0%+	14.4%

* For these grades, 8 and 9, an adaptation of the TEU has been used, cutting the number of questions down from 50 to 25.

the school attended was one in which special economics material had been introduced throughout social studies over a period of three years. Unfortunately, for an unambiguous interpretation of results, students in the experimental schools with special materials were also of higher socioeconomic background so that the effect of the special materials cannot be clearly perceived. Nevertheless, the results are interesting. Taking the course in economics made a small but not significant difference to TEU score. The mean for 31 students who had studied the subject is 25.16; for 177 students who had not, 23.90. Details of the analysis of variance show that only the composite school variable is of value in explaining test results. It may be tentatively concluded that the instruction in economics showed little effect on achievement as measured here but that the special curriculum coupled with differential socioeconomic status made a great difference.[9]

The most intensive work on the development of a semester's economics course for grade 12 has been carried on in Contra Costa County, California, by Suzanne Wiggins from 1964 to date [5]. While the work includes continuous evaluation of both knowledge and attitude with a variety of test instruments, our interest here centers on the use of the TEU because this test has been adopted by virtually every DEEP project.

Over one thousand students in 36 sections taught by 18 teachers were given the TEU before and after instruction in economics. Experimental sections were taught the special "Econ 12" course; control sections received either the regular economics course or no economics course. The results of testing appear in Table 6.

The report [2] from which the table is excerpted is one of the most thorough pieces of research from DEEP and contains important inferences for other studies. Corroborating results, not reproduced here, show breakdowns of test results by teacher, sex of student, education and occupation of head of household, and student ability and achievement. Therefore, I quote a few passages of interpretation from the report:

> A cursory glance [at the table] spells gloom for ECON 12. Control group C shows significantly more improvement in scores on the TEU . . . than any of the ECON 12 groups. . . . the TEU is a good test in the sense that, taken as a whole, it includes questions covering the policy areas and general principles of economics which professional economists would consider important for laymen to know. . . . As such it [TEU] does not test the current content of [the units so far completed of] ECON 12 . . . ECON 12 is not meant to be an introductory survey of economics; it is an introduction to economic thinking.

Conclusion

DEEP has brought about a working partnership among school systems, universities, regional councils for economic education, and the Joint Council. It has succeeded in engaging professional economists in an effort to advance economics instruction through the education of teachers and the development of textual materials—all for the primary purpose of improving students' knowledge.

Measured results of the number of teachers enrolled in economics courses show substantial progress with some form of instruction reaching over 14,000 teachers—about a quarter of the teachers in a typical school system. Sample test results show appreciable gains in teachers' knowledge of the subject.

Measured results of the impact of instruction on students' knowledge are fragmentary and, with few exceptions, inconclusive. It may fairly be claimed that several school systems now have reliable bench marks against which progress may be measured and that sensitivity has been enhanced to the need for controlled evaluation of the instructional process. On this crucial matter I suggest that the amount

---

[9] As pointed out by the Director of Research in the Tulsa public schools, the need for a control on student ability or motivation is suggested in the following table showing grade 12 results for nine classes of representative students who were enrolled and had nearly completed a course in modern history or a course in economics:

	Number of Students	Raw Score (Form B, TEU)	
		Mean	Standard Error
Economic students . . .	135	25.60	.64
Modern history students	123	29.81	.57
	258	27.61	.45

TABLE 6

TEST RESULTS BEFORE AND AFTER INSTRUCTION IN ECON 12, REGULAR ECONOMICS, AND NO ECONOMICS, CONTRA COSTA, 1967

Group	Test Form	N	Scaled Score*		Kr#20†
			X	G	
A: Econ 12	A (Pre)	245	14.7	4.4	.80
	B (Post)	211	16.3	5.3	.83
	Difference of the Mean		1.6		
B: Econ 12	A (Pre)	68	16.7	3.3	.64
	B (Post)	70	19.4	5.4	.84
	Difference of the Mean		2.7		
C: Other economics sections	A (Pre)	81	16.3	4.0	.76
	B (Post)	87	21.2	5.7	.89
	Difference of the Mean		4.9		
D: Other economics sections	A (none)	—	—	—	—
	B (Post)	428	19.2	5.7	.88
E: Econ 12	A (Pre)	187	14.8	3.9	.74
	B (Post)	161	14.9	5.4	.83
	Difference of the Mean		.1		
F: No economics	A (Pre)	158	15.5	4.3	.80
	B (Post)	125	15.6	5.3	.83
	Difference of the Mean		.1		

* Mean scores have been scaled to make scores on Form A and B comparable.

† This is the Kuder-Richardson formula 20 test of internal validity.

NOTE: These statistics should be above .6 to show that students respond normally to any randomly chosen group of test items.

and rigor of evaluation will be directly proportional to the amount of participation in economic education by professional economists.

REFERENCES

1. G. L. BACH AND PHILLIP SAUNDERS, "Economic Education: Aspirations and Achievements," *A.E.R.*, June, 1965, pp. 329–56.
2. *DEEP, 1964–1967* (Office of the Superintendent of Schools of Contra Costa County, Pleasant Hill, Calif., 1967).
3. CAMPBELL R. MCCONNELL, "An Experiment with Television in the Elementary Course," *A.E.R.*, May, 1968, pp. 469–82.
4. M. EUGENE MOYER AND DONALD W. PADEN, "The Relative Effectiveness of Three Methods of Teaching Principles of Economics" (as yet unpublished).
5. SUZANNE E. WIGGINS AND JOHN G. SPERLING, *Final Report, Project H-153, ECON 12* (to U.S. Dept. of Health, Education and Welfare, San Jose, Calif., May, 1968).

DISCUSSION

Lloyd G. Reynolds: One's first reaction on reading these interesting papers is to wonder why this kind of research was not being done fifty years ago. As a profession, we are devoted to research; but our teaching activities have been exempt. We seem to follow a simple labor theory of value: since we put much effort into teaching, there must be a corresponding output in student understanding. It may be psychologically necessary for skilled craftsmen to feel this confidence in their product, but a confidence so easily shaken by simple test results is not very solid.

John Maher's paper suggests, among other things, that we need more information about what our students know on entering college. We assume that they know very little and that anything they think they know is probably wrong because they learned it from high school teachers who were innocent of economics. But this assumption is not as safe as it may have been a generation ago. As the efforts of the Joint Council on Economic Education become more effective and as economics penetrates the curriculum of teacher training nstitutions, we shall get more and more students who know what a trade union and a business corporation are and who have some sophistication in the structure of the economy. This may enable us to abbreviate the institutional background material in the elementary course and get more quickly into the meat of the subject, thus saving some precious classroom hours.

Save them for what? Presumably for exploring analytical tools and applying these tools to descriptive and policy questions. But how well are we doing on this front? I found discouraging, though perhaps not surprising, the Welsh-Fels finding that after a year of college economics a high proportion of students still do not understand the meaning of a demand curve or the simple workings of the price mechanism; and that on some points they seem to have retrogressed compared with what they knew as high school seniors. These conclusions are disturbing enough to call for ntensified research, using instruments of the TUCE type. We need to know much more about variations in student performance by type of institution, size of class, teaching method, level of student ability, and other relevant variables.

It is not hard to surmise what may be happening in our teaching of elementary concepts. We come out of graduate school properly impressed with the subtleties of economic analysis, with the qualifications which must be made to any simplistic statement. Many of us make our early reputations by bending old curves in new directions. The urge to impress these complications on our beginning students is almost irresistible. Such things as the meaning of demand and supply schedules seem too elementary to deserve much attention. I do not think this is just pedagogical laziness. Rather, because we live with these concepts year in and year out, we come genuinely to believe that they are self-evident and need only be mentioned to be understood. The test results described in these papers should disillusion us once and for all on this score.

It appears that in our teaching of economic theory we are simultaneously attempting too much and achieving too little. We are often teaching a mixture of elementary economics, intermediate theory, and first-year graduate theory. The student comes out of this impressed that economics is a hard subject and that his instructor is a bright fellow; but he doesn't have a clear idea of what has been going on. We need to teach a narrower range of concepts, but really to teach them, almost in drillmaster fashion. And we should remember that using economic tools to interpret events in the economy and to analyze policy issues helps to clarify the meaning of the tools.

The Attiyeh-Bach-Lumsden paper is refreshing because it calls into question the accepted folklore of the elementary course. We are all familiar with the conventional wisdom: Economics is a subject which requires discussion, presumably led by really wise people like ourselves. The section method is better than the mass-lecture method. Small sections are superior to large ones. The small, high-quality liberal arts college provides a better learning environment than the large state university. And so on and on.

The Attiyeh-Bach-Lumsden paper reminds us that we don't really know whether these things are true and that we should try to find out. Many of them may not be true. Coming from a "prestige institution," I was really shaken to discover that going to such an institution can have adverse effects on learning. But if the next ten regressions confirm this result, I shall accept it manfully.

The tests of the efficiency of programmed learning are interesting and suggest a need for further experimentation and research. As we all know, there are several strands in the elementary course: (*a*) economic analysis proper; (*b*) using economic analysis to interpret economic events, such as long-term changes in the composition of national product or the characteristics of short-term fluctuations; (*c*) policy applications, where economic analysis is used to lay out the logical structure of a problem and to explore the consequences of alternative actions. One test of success in the elementary course is whether, at the end of it, the student approaches policy questions a bit more like an economist, a bit less like a "layman."

The usefulness of programmed instruction probably lies largely in the first area. In economic theory, definitions are correct or they are not; answers can be clearly right or wrong. This facilitates self-instruction and self-testing, just as it does in accounting or mathematics. When one comes to description and policy, however, answers are less clearly right or wrong. They may take the form, not of "Yes" or "No" but of "It depends on. . ." or "On the one hand . . . and on the other hand." Here I think group discussion and the skills of an experienced instructor have educational value.

In theory proper, the Attiyeh-Bach-Lumsden results suggest that on the average students can learn almost as much on their own with a programmed text as they can in conventional class meetings. But it would be useful to know more about the variance around the average. The comparative advantage of programmed instruction must vary considerably depending on the alternative with which it is being compared. Is the alternative an inferior conventional text plus a mass-lecture or plus sections led by inexperienced teachers? Here programmed learning might actually be superior. Or is the alternative a superior text plus medium-sized discussion sections led by experienced teachers? Here I suspect a good teacher can contribute substantially by clarifying definitions, explaining why they are preferable to other possible definitions, clearing up certain attractive confusions into which students invariably fall, helping to trace complex chains of reasoning, and sustaining student motivation by frequent reference to descriptive and policy applications.

Student motivation, as we all know, is not to be underrated. Over my years at Yale, we have taught the elementary course in almost every conceivable manner. Every few years we are able to get one or more of the "big guns" of the department to lecture for one or two hours a week. It is doubtful that the students hear anything in these lectures which they could not have heard or read elsewhere. But they nevertheless like the format, on the ground that "someone is really paying attention to us," and this may lead them to read more and learn more than they would otherwise.

I have the impression—based on quite limited experience—that even when students find programmed learning effective, they do not find it very interesting. They do not find it the psychic equivalent of lively class discussions led by a good instructor. So while I expect that programmed material will settle into the literature as a useful teaching supplement, it will rarely supersede the "conventional" class meeting.

To return to my first comment, the main moral is that we need more information about the impact of alternative educational techniques. We need improved tests of economic understanding, applied to samples large enough to permit exhaustive statistical analysis. The authors of these papers have done a signal service by opening so many avenues for future work.

Peter O. Steiner: Before commenting on the messages, some bitter, some sweet, that these papers put to us, I feel I must alert you to a danger and enlist your support in defense of what we, as professors, must hold dear. These are times of experiment and change. But defend the basic verities we must. Whatever these studies—and others yet to come—may show, let us not allow programmed learning or the computer or automation or students themselves to deprive the student of the one crucial ingredient of his sound education. Join me if you will in defending to the death that last great bastion of a sound education: the elementary textbook.

I find more comfort in the Maher, Welsh-Fels and the Attiyeh-Bach-Lumsden papers taken together than in them separately, as I shall suggest presently.

The Welsh-Fels paper taken straight reminds me of nothing so much as M. M. Knight's classic definition of education as the process of casting artificial pearls before genuine swine. I am not too disturbed by the fact that there is some unlearning—some would say unteaching—demonstrated by some TUCE questions, because we all know that some confusion is a by-product of increasing sophistication. The fact that poor students tend to know less at the end of the course than good students do at the beginning and the relatively modest improvement in test scores are potentially more troublesome. But I am not beyond hoping that we may be helping students to think constructively about the central economic issues even as we fail to nail down basic principles. It is not at all clear to me that the college teachers' choice is the bleak one these authors suggest between "drawing curves" and "teaching fundamentals," if by fundamentals they mean such things as national income accounting, multiple expansion of credit, and the definition of real wages. We are all concerned to discover the insights of economics, and our courses should strive to share this experience with our students.

The Attiyeh-Bach-Lumsden study offers a number of more encouraging straws: evidently the average student does learn something no matter what we do; programmed learning has a positive marginal product and so do instructors and textbooks. While I find gratifying in principle the fact that some textbooks help more than others, I refrain from asking which one helps most.

Much of the interest in the Attiyeh-Bach-Lumsden paper will center on the intriguing findings of the regression analysis, and while I am tempted to offer my own guesses about the causal factors at work, the fact is that I have nothing sensible to add to the comments made about them in the paper. But all is not roses. What may prove discouraging to some is the same echo in this paper we heard in the Welsh-Fels paper which suggests that we are not achieving high productivity from our elementary courses, at least using the test scores as the measure of product.

I shall refrain from substantial comment on Maher's paper because I am largely ignorant about DEEP and I do not understand what it is that Maher's paper is telling us about the overall impact of that project, but there is nothing there that has changed my impression that economics in the high schools has not proven even a lesser triumph of our profession.

Taken all together, it is clear that these tests and studies tell us something. The problem is, what. There are those who will attack the nature of TUCE and TEU on the ground that multiple-choice questions inherently favor received truth rather than original thought. I think this is probably true. (I remember still my Navy teaching experience when we novice instructors were told that the way to handle midshipmen's questions was to tell them, "Don't ask why, ask how." Students in the sections of instructors who observed this advice performed significantly better on tests than those who did not.) But for all their inherent limitations, I believe these tests are carefully constructed and pretested and they do have a place in reviewing and measuring recognition and understanding of basic principles. What then may we conclude from the poor performance on them?

What I think these studies do show us is that: (1) students do not learn "received principles" easily or accurately in our elementary courses; (2) programmed learning is a promising though not exciting means of teaching them; and (3) the sophistication we inject in our texts and our lectures all too often comes to naught as it finds an inadequate foundation to support it. While I suspect most of us can agree about that, I suspect we will disagree about the appropriate remedy.

The implied solution of Attiyeh-Bach-Lumsden is to use a programmed learning text along with the basic text in the early part of the course in order both to provide a more solid foundation and to free time for the application of theory to social policy. This is a tidy solution, but I believe it is a shade too simple. This is because I feel there is an alternative barrier to clear thinking about the economics of social problems, alternative to not knowing basic economic principles. The second danger is the danger of knowing them.

If we are honest with ourselves and our students, we will admit that what we have to sell is an approach to problems and some hypotheses rather than some neat packages of answers to hard questions. It is after all a subset of the few who do well in our courses and tests that come to us as graduate students "knowing" such things as that everyone is better off under an income tax than a sales tax, that tariffs are stupid, that corporate income taxes do not affect resource allocation, that prices are lower under competition than oligopoly, or that to say that a banking system which keeps 20 percent reserves will add $5.00 to its loans and investments for every $1.00 additionally deposited is false for exactly three reasons.

I welcome programmed learning as a teaching device and objective tests as teaching devices as well as measuring rods, but let us remember that the vital link between our "basic principles" and our "applications to social problems" is the fragile one of sophistication and approach. It is this, a way of approaching problems—a window on the world—that is our stock in trade and that we ought to sell to our students to equip them for citizenship or further study. I doubt that programmed learning helps much here.

Let me propose for discussion, only half in jest, a happy trade based upon the principle of comparative advantage. Evidently we college teachers do badly on very basic simple things, and that here programmed learning has a comparative advantage. At the same time many of us shudder at the thought of high school social studies teachers with their lack of training, "covering the policy areas" or teaching "economic thinking." So let us put programmed learning and "basic principles" into the high schools as a prerequisite to college economics. High school social studies teachers need very little special training to teach this, and I trust them to do it at least as well as their colleagues teach elementary mathematics. Let us, as men and women of sophistication, breadth, training, and vision, welcome students so prepared into our classrooms and try to share with them our insights into why it is that hard questions are hard and easy answers so often foolish. An enormous by-product of this is that much of what students find to be dull and unreal would then be competing with dull geography and dull history, instead of with exciting sociology and political science.

I realize this does not take us very far. We face many problems in the teaching of economic understanding. Among them are dull or unmotivated students and poor and badly trained teachers, to say nothing of texts that are either too dreary or too cute. But there are also good students, good teachers, and exciting issues that need as much talent and understanding as we can bring to bear. While we must measure the incidence of our failures and forge remedial tools, let us not let these tools and the remedial problems blind us to the real challenges that face us and our students.

CAREY C. THOMPSON: Economic literacy and the effectiveness of formal instruction designed to increase its extent have for several decades been of substantial concern to professional economists. In the past twenty years this interest has been, in contrast with earlier periods, actively shared and tangibly supported by many individuals and well-established groups outside the academic world. With this encouragement and support there is perhaps a more hopeful prospect that a knowledge of basic economic principles can be extended to many more young people, that the instruments for such an accomplishment can be improved, and that we can measure the nature and degree of this improvement in the knowledge of economics. Many academic economists, including some of the most distinguished members of the profession, have been and are involved quite actively in the efforts of recent years. They have faced many very difficult tasks, and their efforts merit the commendation of their colleagues. In calling attention, as I subsequently do, to a few hazards in these undertakings, I wish to make it clear that my comments are not meant to attribute motives of questionable character to anyone involved in the economic-education-improvement effort. I would also assume that my warnings, if such they could be called, have occurred to most, if not all, who have labored in this field.

As I read the results of the studies presented at this session, it seems fair to draw the following tentative conclusions: (1) economics instruction at the basic level in college at present is far from successful in getting students to understand what we, generally, have expected them to learn; (2) some increase in the effectiveness of that instruction can be achieved by the use of programmed learning methods; and (3) a faint hope of improvement in the efficiency of economic education can be seen in the extension of efforts to this end now being carried on in precollege education.

These conclusions yield little, if any, encouragement to those of us in the profession who are concerned about the state and the prospects of instruction in basic economics. However, those of us who have been long involved in such instruction or who have surveyed earlier concern or efforts at improvement will not, I think, be surprised at the current conclusions.

But whatever our reactions, it seems safe to say that much remains to be done before any of those concerned, economists or noneconomists, can be reasonably well satisfied with our accomplishments in economic education. So, what about the future?

It would be my guess that, given a few years and some very substantial effort, we could raise appreciably the very low percentage of students who, exposed for a half-year, say, to elementary analysis, could distinguish between demand and quantity demanded. We might also get quite a bit better reaction—reflex, that is—to the process of deposit-creation.

But surely we might wish, even in basic economics, to try for student reaction and understanding that would go beyond the numbers, or the curves. Just what those additional steps might be would become harder to define or to agree upon, yet if basic principles have any significance to the young or old citizen-economist, that meaning will have substance only as these principles are put to some use, such as passing judgment on policies used (or suggested), rules under which the economy does (or might) operate, and so on. Some of us might think that the most efficiently trained of the students who could go on beyond the numbers would be those most interested and willing to provide a challenge to policies, rules, and arrangements currently in existence—even those in the local community itself.

Is this, I wonder, the kind of economic literacy toward which we strive—if we take a longer-range point of view? I have a feeling that, as any such goal as this were moved toward, the enthusiasm among many current supporters of the drive would diminish very greatly indeed, particularly if the accomplishment could be extended into the secondary school stage.

Perhaps I err in viewing economic literacy as an attainment that would produce and encourage reevaluation of the existing rules and the results thereof, especially in terms of the distribution of income and wealth. I cannot, however, see it as a completely innocent and innocuous intellectual accomplishment—something to be achieved but not to be used. I strongly suspect that much of the motivation for those who have it as a desirable goal is the conviction that economic literacy would result in added support for "free enterprise," as the latter is conceived by its advocates. It is, indeed, difficult to escape the conclusion that much of the widespread interest in economic education from outside the ranks of professional economists comes out of "cold war" influences, with the implication that economic analysis should carry a heavy overburden of indoctrination against the errors of distasteful alien doctrines and practices.

TRANSPORTATION AND THE PUBLIC UTILITIES

TRANSPORTATION REGULATION AND ECONOMIC EFFICIENCY

By MARTIN T. FARRIS
Arizona State University

At the outset it must be noted that this subject is far from being new. The monopoly question was one of the major topics of the first meeting of the American Economic Association in 1886 and "The Relation of the State to Industrial Action," by Henry C. Adams, is one of the papers appearing in Volume I of the American Economic Association publication ("First Series"). It is fairly obvious that economists and students of transportation have been interested in the effects of regulation on efficiency for a very long time.

As long as society continues to regulate, the problem will continue to exist. Conditions gradually change and regulation itself gradually changes so the problem keeps coming back again and again for consideration. One might also add that the economists' views of regulation and efficiency also are subject to change.

The purpose of this paper is to suggest that our traditional ideas about regulation and its effects on efficiency need to be reexamined once more. Such reexamination may point the way to some modification or change of our views on regulation in response to the slow changes taking place in the environment in which regulation operates and in response to the changing views of economics itself.

Transportation Regulation and Monopoly

Transportation regulation has been around for a long, long time. While regulation at the federal level is now over eighty years old, regulation by states via the statutes goes back even further. If one examines the roots of the common law regulation of the common carrier by way of the concept of the "common calling," regulation extends back to at least Lord Hale in 1670. If one seeks the foundation of liability law, it is found that the "public" was imposing special obligations on the carriers in ancient times.

Early regulation was based on the ideas of "monopoly," of course. Sometimes the monopolist was a creature of the sovereign or the state itself when certain persons were given exclusive rights to a "common calling." In return for this monopolistic position, the recipient of such a right or franchise was regulated "in the public interest."

Modern transportation regulation beginning with *Munn* v. *Illinois* (1876) and the Act to Regulate Commerce (1887) is also based on monopoly. While "entry control" was not the dominant form of regulation, the economies of scale inherent in the enterprises "affected by the public interest" made them monopolistic in form. And, indeed, the railroads—the first mode of transportation regulated extensively by statute—were monopolies. Entry was in fact difficult, competition was limited at all but terminal points, and the public was exploited in many cases.

Regulation under such circumstances was considered about the only workable substitute for competition. The public ownership alternative to regulation in order to protect the public was sometimes proposed but was generally rejected as being foreign to the "American philosophy of free enterprise." Therefore, given the prevailing economic philosophy of the times and the almost complete reliance on private ownership, regulation was the only alternative.

Further, if competition led to economic efficiency (another prevailing belief) and regulation was a substitute for competition, then regulation led to economic efficiency. If this argument was too tenuous to be accepted, the justification of regulation as the only "practical way" to "protect" the public could always be brought forth.

There is a long history of the perfection of the regulation of transportation monopoly between the original federal act in 1887 and the imposition of regulation on other transportation modes principally in the 1930's and 1940's. To be sure, petroleum pipelines were regulated in 1906 along with various accessorial rail services, but it was the rail monopoly that was the principal concern of regulation. By the Act of 1920, the regulation of monopoly was essentially perfected although a few amendments were added to deal with the labor and the bankruptcy problems in 1926 and 1933. Thus, by the early 1930's the "monopoly pattern of regulation" was completely developed as an integral portion of national transportation policy.

Under the monopoly pattern, regulation was assigned the role of substituting for competition. While there could be some argument about economic efficiency, there seemed to be very little choice since competition was not attainable given the economic characteristics of monopoly.

This policy, however, was a very static one. In the meantime, the transportation market was slowly evolving away from the monopolistic pattern. The highway system was developing rapidly under the twin supports of federal aid and the state gasoline tax. Highways provided the "way" for the trucklines and the competition of the 1950's and 1960's was born. Likewise river improvements and developments plus generous federal assistance to the infant airline industry were spawning more competition. Finally, technological developments during the 1940's were to lead to a much more viable and strong pipeline industry. The end of monopoly by railroads in transportation was in sight.

The pity of the matter was that transportation regulatory policy was so slow to recognize the changed market conditions. Indeed, the monopoly pattern was imposed upon the trucklines in 1935, the air carriers in 1938 (with modifications), the inland water carriers in 1940, and the freight forwarders in 1942. The final use of the philosophy of monopoly was the legalization of the Rate Bureaus in 1948 with the amendments contained in the Reed-Bulwinkle Act.

All of these policy pronouncements were in the monopoly pattern and had as their justification protection of the public from monopoly—in spite of the fact that transportation monopoly was a thing of the past. Where questions were posed as to the dissimilarity of the economic characteristics between railroads and the newly regulated transportation modes, the "equity" argument was used to justify imposing the "monopoly pattern." Briefly stated, this equity argument contended that since the modes competed for the same business, it was only equitable that both be regulated in the same way. The upshot of the protection of the public and the equity argument was that the pattern of railroad regulation was imposed on all modes of transportation even though they did not possess the same economic characteristics and were not all monopolies.

Competition in Transportation

As students of transportation and economists, the academic community pointed out these inconsistencies long ago. General dissatisfaction with regulation and the recommendation of "more competition" also has a long history. Generally it was felt that the market mechanism was better able to allocate resources and to judge the action of the particular transportation mode. Although scholars were concerned with the problem of competition among the rail giants where it existed, they were more concerned over the lack of competition where the monopolist was the only carrier. Here regulation was the only answer.

Yet economists, being imbued with the idea that the market mechanism was the best allocator and the only route to economic efficiency, really never accepted regulation as the answer. At best, regulation was a poor second choice to competition in the mind of the economist.

Further, when "intermodal competition" became more evident as substitute modes of transportation arose in the 1930's, regulation became even more suspect. Where transportation modes competed for the same business (freight), price became very important. But prices were determined by regulation, by an administrative body charged with "protecting the public." It was soon apparent that regulation was assuming a new role: that of allocating the traffic between modes.

Evidence of this new role of allocation is well known. The regulatory bodies found themselves more and more faced with the problem of setting minimum rates among competitors and less concerned with holding down the maximum rates of the monopolist. The main regulatory concern shifted from protecting the public from possible monopolistic exploitation to concern with protecting one mode from the competitive aggressiveness of a competing mode. Literally the ICC not only became the allocator of the traffic but also the protector of the weaker modes. And Congress formalized this role by declaring it to be the "National Transportation Policy" in 1940 to "recognize and preserve the inherent advantage of each [mode]."

At this juncture, scholars became even more concerned with the question of regulation and transportation efficiency. Not only was regulation costly in itself, but it was human and subject to errors. In the hands of various regulatory authorities with long histories of monopoly thinking, regulation tended to become highly institutionalized. Indeed, it became popular for economists to concern themselves with the problems and mistakes of regulation, the background of regulators, and the "life cycle" of regulatory institutions.

The obvious solution was always recommended: return to competition and let the market play a bigger role. This "solution" reached its greatest popularity when President John F. Kennedy sent a special "Transportation Message" to Congress in April, 1962. This special message was the first time that transportation regulation had been so singled out for attention and, more importantly to the economist, the message called for a return to competition. Kennedy made a plea for more competition and the reliance on the natural forces of the market rather than reliance on administrative judgments in regulation. It appeared that at long last the economists' plea for more competition as an avenue to economic efficiency was being heard.

As is well known, Congress did not act decisively. Some bills were introduced, some hearings were held, but no startling change in public policy came about. The vested interests of the various carriers and the regulators coupled with real concern over just what would happen if more competition were injected into transportation led to delay and finally defeat for meaningful action. Economists had their hero; textbook writers could thereafter cite the President's message as final authority to substantiate what had become the byword of all economists: We need competition. But no change seemed likely to take place.

An Oligopolistic Market?

Six and a half years have passed since the high point of the plea for competition took place. The time has now come to reassess the situation and to be sure that the plea for competition continues to be a sound policy recommendation and not a "conventional wisdom" among aging gladiators. A careful look at the assumptions of the argument and the conditions of the case is called for.

Taking an aggregate view of the structure of the transportation market, the number of firms in all modes except motor is certainly small. In air, the number is well under thirty in the first two levels of the market (the trunklines and the feeder lines), in rail under eighty (in the Class I group), and in pipeline and inland water transportation, under one hundred. The aggregate view could not lead to the conclusion of monopoly, but it is certainly not anything approaching competition either (as the economist uses that term). The transportation industry appears to be somewhat oligopolistic.

From a disaggregated point of view, the tendency toward oligopoly becomes even clearer. Looking at the choice of carriers of a given mode on a given route, it is readily apparent that only a very few alternatives are available to the shipper. On any given movement (discounting very circuitous routings), only a very small number of transportation firms compete for a given amount of traffic. Something which closely approaches the classical "small numbers case" prevails.

When entry controls via certificates of public convenience and necessity are added, even the motor carrier industry appears oligopolistic. Even on the routes with the highest density, only a handful of motor carriers compete with one another for traffic. Truly an atomistic view of the structure of transportation discloses a high degree of oligopoly in every mode.

Of course, this bit of information is not new. Indeed this was the basis for the original monopolistic concern with the railroads in the first place. The economies of scale apparently are such that only a few firms can profitably operate on any given route. The possible exception to this might be in the motor carrier field where entry control, not economies of scale, seem to enforce the oligopolistic pattern.

Turning to behavior patterns, all firms of a given mode on a given route are highly interdependent and interrelated, with the action of any one carrier drastically affecting the business of any other carrier. Examples of this are common knowledge. This interdependence may be in terms of equipment (example: air transportation and the need for all firms to "keep up" in the introduction of new jet planes), sales campaigns (when one railroad sets up a "marketing department," all rush to serve their customers with such departments), credit extensions (the problem of collection of bills in the motor carrier business is common knowledge—in spite of the credit extension regulations), "free" services (railroads compete in ability to "trace"), and in many other ways.

Finally, "service" is a most tenuous and indefinable factor in the transportation market. It may depend almost entirely on "what a nice guy" the traffic solicitor turns out to be or which salesman treats the shipper the best. Here is a classic example of the firms competing on the basis of "imagined differentiations in product (service)"!

Intermodal Competition and Oligopoly

But all of this is not new. Economists have long recognized "competition among the few" by mode. Further, recognition has been given at least since the 1930's to the idea that the various modes of transportation are substitutes for one another in a substantially relevant range. This is what the term "intermodal competition" is all about.

But the misleading word here has been "competition." Economists have not really meant competition in the traditional sense of many sellers competing with similar products (service). This use of the term "competition" is possible only if homogeneity of service ("ton miles"?) is assumed as Taussig erroneously did so many years ago in his classic transportation debate with Pigou.

Actually, each mode has a range of distinctive services and a range of competitive services. Degrees of substitution exist. The situation is very analogous to the oligopolistic situation where a Volkswagon competes with a Cadillac. Each is distinctive and has its own market attributes, yet each competes, too. However, the relevant range of competition or substitutability is small in this example. The relevant range of substitutability and competition increases when Chevrolets and Fords are considered. The same holds for transportation. Railroads and airlines compete in the hauling of freight but the relevant range of substitutability and competition is small. Railroads and trucklines compete in the same market and the relevant range of substitutability and competition is large.

Economists have observed this phenomenon in the general economy for some time and have discussed it in terms of oligopolies. Triffin had a term for a market with few firms with distinctive products: heteropolies. When the degree of interdependency was high, he added the word "circular" since no one firm could change in any way without affecting its companion oligopolist. Transportation today comes very close to Triffin's "circular heteropoly," and this is really what is meant by the term "intermodal competition."

Again, the oligopolistic view of the disaggregated transportation market is not entirely new. Meyer, Peck, Stenason and Zwick forcefully pointed out these structural characteristics in 1959. What is interesting, however, has been the traditional transportation economists' hesitancy to go further than merely pointing out these oligopolistic factors. Rather than developing on these apparent examples of oligopoly and following modern oligopoly theory and discussions, the transportation economist has been prone to retreat back to the "safe" plea for "more competition."

The reason for this lack of application of oligopoly theory has perhaps been twofold. On the one hand, transportation economists have been steeped in the tradition of competition. While the criticisms of the competitive model were well known, it still remained the safe position, the traditional position. Perhaps it has been the easy position as well: one could always be in favor of more competition. Even in the face of the knowledge that competition probably wouldn't work (at least where economies of scale and entry barriers in the form of mammoth investment were found), the transportation economist still felt somehow obligated to defend competition as the route to economic efficiency.

The second reason for taking the safe position of more competition was because modern oligopoly theory (at least implicitly) rested upon the firm's ability to compete freely among the few. That is, it assumes a lack of public regulation of entry, public regulation of price, and public regulation of service—and all knew that in transportation these very factors of entry, price, and service were indeed regulated.

The Role of Regulation and Transportation Oligopoly

This latter rationalization of the "more competition" syndrome is founded on a basic misunderstanding of the role of transportation regulation relative to oligopoly. As long as regulation was thought of solely as a substitute for competition, the connection between regulation and oligopoly was overlooked.

Actually transportation regulation enforces oligopoly on the one hand and delays the market responses of "competition among the few" on the other. By way of entry controls, oligopoly on any given route (market) is established and enforced. The small number of carriers and the resultant interdependency is a direct result of entry regulation in motor, air, inland water, and freight forwarder modes. The case is less clear in rail transportation, but even here entry and exit controls affect changes in the rail network and hence perpetrate the oligopolistic rail pattern.

By causing price (rate) changes to be made slowly and by way of a highly institutionalized process, transportation regulation merely delays the natural oligopolistic responses to interdependency when "competition among the few" prevails. Certainly no one doubts that the railroads would "meet and beat" all rate cuts proposed by the trucklines if they were not restrained by regulation? The regulatory requirement of published rates, docketing proposals, hearings and opportunity for protests, and many other regulatory devices designed to protect the public are, in effect, instruments of delay. "Suspension and Investigation" procedure is, of course, the prime example of this "delay of natural oligopolistic responses."

Properly understood, transportation regulation takes on an added role: the enforcer of oligopoly and the delayer of natural oligopolistic responses.

Conclusions

There are two implications to the idea that transportation is indeed an oligopoly and regulation merely enforces this market classification and delays the natural market responses. First, economists need to redirect their view more towards the role of regulation and oligopoly rather than pleading for a competitive market. Recognition can be given to the fact that transportation regulation is probably here to stay and rather than accepting it as a "case of institutionalized frustration" (the catchy phrase developed by Sidney Miller, Jr.), recommendations to improve regulation should be developed. But those suggestions for improvement must be innovative and based on recognition that oligopoly theory applies. They cannot be the same old pleas for merely "more competition." As Miller stated in 1961, there must be "the courage to eradicate institutionalized frustration."

Second, economists need to examine the internal workings of transportation regulation using oligopoly theory. Why not apply "kinky demand curve analysis," price leadership by dominant firms or "barometric firms," and the theory of "joint profit maximization" to transportation? Performance testing, behavior patterns, and market structure tests have all been thought out and proposed by the student of oligopoly. Don't they also apply to transportation?

As one example, the Section 5a Rate Bureau offers a very fruitful area for the application of oligopoly theory. Most certainly "bargaining" takes place. Some, if not all, of the "four factors pertaining to the outcome within the bargaining range" as developed by Fellner certainly apply. In the Section 5a context, how important are "(*a*) long-run consequences of violating accepted value judgments," or "(*b*) the immediate political consequences of a stalemate," or "(*c*) the ability of the parties to take and to inflict losses during stalemates," or "(*d*) toughness in the sense of unwillingness to yield in a range in which the other party is expected to yield if one fails to do so"? Or another example, what evidence is found of "improvement" of oligopoly by "nonprice competition" or by "cost and product (service) differences"? What are the "safety margin considerations"?

The point is, economists have not adequately considered the workings of regulation and how it might be improved, explained, or understood. Rather there has been almost total concern with only the end result of regulation. Modern oligopoly theory has not been applied and transportation economists have clung to the neoclassical competitive model as a *modus operandi*. Perhaps it is time for transportation economists to update their approach and their thinking about transportation regulation.

CONGESTION THEORY AND TRANSPORT INVESTMENT

By William S. Vickrey
Columbia University

"Deepening" and "Widening" of Transport Investment

Investment in transport facilities necessarily begins by being largely investment in the provision of new routes or new services under conditions of substantial indivisibilities and increasing returns to scale. Under these conditions the usual profitability tests for determining the desirability of specific investments lead generally to under- rather than to over-investment in transportation facilities. At this stage, cost-benefit analysis needs to include substantial elements of consumers' surplus on the benefit side in order to arrive at correct evaluations.

As investment proceeds, however, larger and larger proportions of transportation investment are made primarily, or at least in large measure, to relieve congestion on existing routes and to expand overall capacity. In such instances criteria based on apparent profitability may be seriously misleading in the opposite direction, and when notions of consumers' surplus are narrowly applied without regard to the overall situation, the errors may be compounded. This is especially likely to be the case where charges levied for the use of the existing and prospective competing facilities are far wide of the mark of representing marginal cost, as they often tend to be. It is this latter type of investment, designed to relieve congestion, with which this paper is concerned.

Types of Congestion

For purposes of economic analysis it is useful to distinguish at least six types of congested situations, though they are in fact often encountered in various combinations. These can be designated simple interaction, multiple interaction, bottleneck, triggerneck, network and control, and general density.

Single interaction occurs whenever two transportation units approach each other closely enough so that one or the other must be delayed in order to reduce the likelihood of a collision, no other units being sufficiently close to be immediately affected. This is the chief form of congestion encountered in light traffic. Total congestion delay tends to vary as the square of the volume of traffic; thus a motorist deciding on a trip under light traffic conditions will thereby inflict on others an amount of additional delay roughly equal to that which he himself will experience. (To be sure, for some types of vehicles the effect may not be symmetrical: slow-moving vehicles may tend to be relatively little delayed and fast-moving vehicles relatively more. The above relationship holds for an average vehicle.)

Multiple interaction tends to take place at higher levels of traffic density, short of capacity flows, where one can expect the average speed s to be a function of the flow of traffic x: $s=f(x)$. For traffic volumes ranging from about 0.5 to 0.9 of capacity, one can often fit a function of the form

$$(1) \qquad z = t - t_0 = \frac{1}{s} - \frac{1}{s_0} = ax^k$$

where t is the time required to go a unit distance under actual conditions, t_0 is the time required under very light traffic conditions, z is the average delay per vehicle, and a and k are constant parameters. For a relationship of this form, the total increment of delay given by

$$\frac{d(zx)}{dx} = z + x\frac{dz}{dx} = ax^k + xak\,x^{k-1} = (1+k)z \tag{2}$$

that results from a unit increment of traffic thus works out to $k+1$ times the delay experienced by the vehicle itself. That is, for every minute of delay directly experienced by the added vehicle, k minutes of delay are inflicted on the remaining traffic. For situations where considerable congestion exists, k is likely to be in the range of from 3 to 5 or even higher. The previous case is essentially that where $k=1$.

The pure bottleneck situation, which is the one that will chiefly concern us here, is one where a relatively short route segment has a fixed capacity substantially smaller relative to traffic demand than that of preceding or succeeding segments. There is thus relatively little delay as long as traffic remains below the capacity of the bottleneck, though small amounts of delay may occur as a result of stochastic variations in the level of traffic flow when the average flow is just below capacity. The important delays will occur when desired traffic flow continuously exceeds the capacity of the bottleneck for substantial periods. We then find that queues accumulate until either a period is reached when traffic demand is below capacity, or the prospect of waiting in the queue reduces the traffic demand by diverting it to another time or route or by suppressing the trip entirely.

A triggerneck situation develops from a bottleneck situation whenever the queue backed up from the bottleneck interferes with the flow of traffic not itself intending to use the bottleneck facility. The onset of incremental congestion may be quite sharp, and indeed in extreme circumstances a circular chain of triggerneck situations may bring traffic to a complete standstill, requiring that at least some of the vehicles involved actually back up before a forward movement can be resumed.

Network and control congestion results whenever the levels of traffic during the peak are reached requiring the application of additional control measures, whether in the form of regulations, stop signs, routing limitations, traffic lights, train controls, flight patterns and rules, or otherwise. Aside from the cost of these measures in themselves and even assuming that they are invoked only when the circumstances without them would be demonstrably worse, either from the standpoint of safety or delays, it is generally true that they cannot be applied with complete selectivity as to time and place, so that in most cases control measures required to take care of the most severe conditions will result in more delay under less severe conditions than would occur in the absence of controls, or with less restrictive controls adapted to the less severe conditions. Thus some of the delay experienced by off-peak traffic is in a medium-long-run sense caused by the increase in the peak traffic that made the controls necessary.

Finally there is a sense in which congestion costs in the long run are a function of the overall density of transportation flows in a given area for all modes combined and over all routes, even though some modes may contribute less to the total overall congestion relative to its traffic volume than other modes. Even if route separations are such that traffic on one mode has no immediate impact on traffic on the other modes, the construction of facilities to accommodate additional traffic on the one mode or route will not only encounter increased construction costs by reason of other existing facilities that cross its path, but such construction will at the same time be increasing the cost of constructing any other transportation facilities across its path in the future. In highly congested areas there is a very real sense in which long-run increasing costs may be encountered. It is very rarely that any account is taken, in the estimating of the cost of constructing facilities for a given transportation route and mode, of the increased costs of such future crossings by other links, even though good technical planning may sometimes make provision for such future crossings in the design.

Accidents as a Cost of Congestion

In addition to the cost of delays, the cost of accidents constitute an often overlooked element in the costs of congestion. While the incidence of traffic accidents does not arise with traffic density quite as rapidly as do time delays, one does expect, a priori, that as vehicle interactions per vehicle-mile increase, accidents per vehicle-mile will also increase. There is, indeed, a certain amount of empirical evidence that in a significantly wide range of situations this increase in accident rates with increasing traffic densities does in fact occur: for grade-separated limited access highways in California, it was found that the marginal increment in the number of accidents associated with an increment of traffic on a given type of highway was approximately 1.5 times the average accident incidence per vehicle-mile.[1] Thus whatever may be the effect on accidents of shifting traffic from other highways to grade-separated expressways, there is in addition a favorable effect on accidents of building roads of the same type to more ample dimensions and greater capacity, and an adverse effect on the accident rate per vehicle-mile of increasing the flow of traffic on a given roadway.

Before taking full credit for this benefit, however, it is necessary to examine the net safety effects of increasing total traffic flow overall, and of attracting traffic from other safer modes, such as rail transit. Doing too much in the name of safety considered in a narrow context can actually increase the overall death rate.

Construction to Ease Bottlenecks

Although the pure bottleneck situation is not typical of the general congestion picture, it is an important element in many cases of severe congestion and its relatively simple analysis does provide some valuable insights into the nature of the overall congestion problem.

Assume a situation in which $N=7200$ commuters want to make a daily trip via a given bottleneck, and that in the absence of congestion their times of passing the bottleneck point would be distributed evenly over a period between $t_a=8{:}00$ A.M. and $t_b=9{:}00$ A.M., this permitting each commuter to arrive at his downtown destination at a desired time. If the capacity of the bottleneck were to be enlarged to

$$v_m = \frac{N}{t_b - t_a} = 120 \text{ cars per minute,}$$

then of course the capacity would just meet the requirements and no queue other than that due to stochastic variation would occur.

If the capacity is kept smaller than this, i.e., $v<v_m$, then it becomes impossible for all the commuters to arrive at their destinations just at the desired times. Some, at least, will have to arrive either late or early. In the absence of tolls the steady state that results will involve varying degrees of queuing, with those arriving at their offices closest to their desired times generally having to spend relatively more time in the queue than those who choose to push their arrival time further away from the desired time.

To keep the model simple, let us suppose that all commuters uniformly value time spent at home at $w_h=2$ cents per minute, and time spent at the office at w_o, which for time prior to the desired starting time we suppose is $w_o=w_p=1$ cent per minute, and for time after the desired starting time is $w_o=w_j=4$ cents per minute. Time spent in the queue has a value of $w_q=0$. It is readily seen that if an individual is to be maximizing the overall value of his time, he must be leaving the bottleneck point, subsequent to any queuing he may have had to endure, at a time such that

---

[1] See William Vickrey, "Automobile Accidents, Tort Law, Externalities, and Insurance," *Law and Contemporary Problems*, Summer, 1968, pp. 467–68.

$$\begin{aligned}\frac{dq}{dt} &= \frac{w_h-w_0}{w_h-w_q} \\ &= \begin{cases} 0.5 & \text{for } w_0=w_p=1 \text{ cent/min} \\ -1.0 & \text{for } w_0=w_j=4 \text{ cents/min}\end{cases}\end{aligned} \tag{3}$$

where $q(t)$ is the amount of waiting in the queue required in order to leave the bottleneck point at time t. A fraction

$$r=\frac{w_j-w_h}{w_j-w_p}=\frac{2}{3} \tag{4}$$

of the commuters will pass the bottleneck during the period of queue buildup and arrive at work at or before the desired starting time, the remaining fraction $1-r=1/3$ of the commuters will leave the bottleneck after 8:40 A.M. during the working off of the queue and arrive at or after the desired time. The total time required for the commuters to pass the bottleneck will be $N/v=7200/v$, which will also be the length of time that a queue will persist, as long as $v<v_m$.

The length of the queue will build up linearly from zero at time

$$\begin{aligned} t_i &= t_a - r[(N/v)-(t_b-t_a)] \\ &= 8{:}40 - 4800/v \end{aligned} \tag{5}$$

to a maximum wait in the queue of

$$q_p=\frac{N}{v}(r)\frac{w_h-w_p}{w_h-w_q}=\frac{2400}{v} \tag{6}$$

for cars leaving the bottleneck at

$$t_p = t_a + r(t_b-t_a) = 8{:}40 \text{ AM} \tag{7}$$

after which it will again decline linearly to zero at

$$\begin{aligned} t_j &= t_b + (1-r)[(N/v)-(t_b-t_a)] \\ &= 8{:}40 + \frac{2400}{v}\cdot \end{aligned} \tag{8}$$

There will be a sharp discontinuity in the amount of delay experienced as the capacity of the facility is expanded past the point where $v=v_m=120$ cars/minute. Below v_m, delay is inversely proportional to the capacity v, while at and above v_m the delay from queuing is zero. To be sure, matters will not usually work out as sharply as this, for there will usually be some variation in the desired rates of traffic flow near the peak rather than a peak that has an absolutely flat top, as has been assumed here for the sake of simplicity. Moreover, there will usually be some elasticity of traffic demand with respect to queuing time such as to suppress some traffic entirely rather than merely to shift its timing. Nevertheless, sharp discontinuities such as this emphasize the need for careful analysis of practical situations.

In practice, too, we are usually dealing with a dynamic situation in which traffic levels are generally growing at a substantial rate, while construction of additional facilities takes time and usually involves substantial lumps of additional capacity. In the face of such substantial penalties for either over- or underinvestment, substantial waste is likely unless some form of control over the use of the facilities is applied, such as is available through pricing, and which does not involve the wastes of queuing. Unlike the construction of additional capacity, prices can be adjusted upward or downward, as proves to be desirable, on relatively short notice and by relatively small increments.

Indeed, in the above situation one can readily compute the price structure that would just eliminate the queue and lead to efficient use, at least in the short run, of whatever facilities are actually in place. This will consist of a toll rising linearly from 0 at t_i to

$$p_p = \frac{Nr}{v}(w_h - w_p) = \frac{4800}{v}, \tag{9}$$

at t_p and then declining linearly to zero at t_j. With this pattern of tolls, each commuter will find that he can do no better for himself than to set out in time to pass the bottleneck at the same time that he would have left it after waiting in the queue in the zero (or constant) price situation. If traffic should fail to adjust its movement in such a fashion as to eliminate the queue, those finding themselves in the queue would have a motive to shift their travel time in such a way that the queue would be eliminated.

In the short run, the commuters are just as well off paying the variable toll and having no queue as they were before with no toll but with an equivalent queue; moreover there is no change in the pattern of arrivals at the city center. The revenue derived from the charges thus represents clear gain. We thus have an example of tax revenue that not only has no excess burden, it has no burden at all! To the extent that any of the revenue from the variable toll is used to reduce a preexisting flat toll, the motorists will be better off.

Obviously this does not mean that expansion of the capacity of the facility is never justified, but it does mean that the justification for such investment must be considered in an entirely different light if congestion charges are a possibility than if they are not. In the absence of congestion charges, a decision to expand facilities may have to be taken on an all or nothing basis. Expansion inadequate to take care of the entire traffic demand may result in a relatively slight improvement in conditions and may turn out to be hardly worthwhile, while a just slightly larger expansion might clear conditions up rather dramatically. Unfortunately, in a dynamically changing situation it is difficult to predict just what size of an improvement will in fact get over this threshold and for how long. It is easy to think of cases where an expansion of capacity felt to be quite ample when planned has turned out to serve merely to attract additional traffic until conditions are almost as bad as they were originally.

If, under conditions similar to the above, the levying of congestion charges is either an actuality or an alternative under consideration, benefits from the expansion of capacity are likely to be both smaller and less capricious in their behavior than if no pricing is contemplated. The net gain from the expansion of the bottleneck, assuming the adjustment of charges both before and after so as to just eliminate queuing, consists not of any reduction in queuing time (since there isn't any queuing) nor still less of the reduction in tolls, since this is merely a transfer from the government or operating agency to the users (and may entail substantial costs involved in securing an equivalent revenue from other sources), but simply in the fact that users will be traveling at times closer to the preferred times. The value of this shift in time may be measured by the difference in the value they place on their time at the two ends of the journey, i.e., $w_h - w_p = 1$ cent/minute for reductions in the amount by which commuters travel in advance of the preferred time, and $w_j - w_h = 2$ cents/minute for reductions in lateness. The total value of this delay, under optimum charging and no queue, is given by

$$
\begin{aligned}
&Nr(t_a - t_a)(w_h - w_p)/2 \\
&\qquad + N(1 - r)(t_\beta - t_b)(w_j - w_h)/2 \\
(10)\quad &= N^2 r \frac{1 - r}{2}(w_j - w_p)\left[\frac{1}{v} - \frac{1}{v_w}\right] \\
&= \$1440\,\frac{120 - v}{v}\,.
\end{aligned}
$$

Where queuing results from the absence of tolls, the average queuing time is half the maximum, or $q_p(\frac{1}{2}) = 1200/v$, which when evaluated for 7,200 cars per day at w_h 2 cents/minute amounts to $\$172,800/v$.

The results, for various values of v, are summarized in Table 1.

Imposition of the optimal variable toll in each case eliminates queuing and results in toll revenues equal to the cost of the eliminated queuing. The displaced arrival cost will be the same whether or not the optimal toll is imposed, as this depends only on the capacity of the facility.

If we now suppose that initially we have a 2-lane bottleneck with $v = 60$ cars per minute, then without the control provided by the variation in the toll rate congestion costs will total $4,320 per day. Without help from toll adjustments, opening up a third lane at a cost of $2,000 per day would reduce congestion costs by only $1,920, and so would not be worthwhile, although opening up two new lanes at a cost of $4,000 would eliminate the entire $4,320 worth of congestion and yield a net gain of $320. On the other hand, the institution of variable tolls according to an appropriate pattern would cut congestion costs of the 2-lane bottleneck by ⅔ to $1,440 and result in a budget inflow of $2,880 instead of a budget outflow of $4,000 if the additional two lanes were built. If the overhead costs of obtaining public funds from other sources were as much as 10 percent—a not unreasonable figure if all of the unfavorable results of increased tax rates are allowed for—this budgetary shift would constitute a further gain of $688, for a total gain of $2,880+$688 =$3,568, a substantially better result than any that can be obtained without toll variation, and much better, in particular, than the $320 gain from the 2-lane expansion.

If, on the other hand, the cost of expanding the bottleneck is relatively low, say $20 per day for each vehicle per minute of increase in capacity, so that some addition to capacity will be worthwhile in any event, variable tolls can still play a

TABLE 1

Capacity (cars per minute)	Equivalent Number of Lanes	Duration of Queue or Toll (minutes)	Maximum Wait in Queue (minutes)	Average Toll Rate (cents)	Congestion Cost ($/day)		
					Displaced Arrival	Waiting in Queue (= toll rev.)	Total without Pricing
50	1.67	144.00	48.00	48.00	2016	3456	5472
60	2.00	120.00	40.00	40.0	1440	2880	4320
70	2.33	102.9	34.29	34.3	1029	2469	3498
80	2.67	90.0	30.00	30.0	720	2160	2880
90	3.00	80.0	26.67	26.7	480	1920	2400
100	3.33	72.0	24.00	24.0	288	1728	2016
110	3.67	65.6	21.91	21.9	131	1571	1708
115	3.83	62.6	20.87	20.9	63	1503	1566
118	3.93	61.0	20.33	20.3	24	1464	1488
119	3.97	60.5	20.17	20.2	12	1452	1464
119.999	4.00−	60.0	20.00	20.0	0.12	1440	1440
120.001	4.00+	0	0	0	0	0	0

role in reaching the optimal result. In the absence of any tolls the best available alternative would again be to expand capacity to slightly above $v=120$ cars per minute, so as to eliminate congestion entirely. If toll controls are available, however, it would not pay to carry the expansion much beyond $v=90$, since expanding from $v=90$ to $v=100$ would reduce losses from displaced arrival by only $192, as compared with the cost of $200 for this expansion. In addition, such an increase might entail an increase in budgetary problems by the difference in revenues of $392.

The use of congestion tolls as an element in developing an efficient transportation system is thus not only a means of providing optimal adjustment in the short run but is likely to remain an important element even in the long run. It is only if the increment to capacity is provided in a manner that incidentally provides a substantially new and different route or a shorter origin to destination time for a substantial amount of traffic, or possibly where incremental costs of adding facilities for the collection of congestion charges *de novo* bulk large, that it would be possible to omit such charges from an optimally efficient scheme.

In practice, of course, most bottleneck situations are not as simple or as clear cut as the above case. The desired times of passing the bottleneck are not usually uniformly distributed, individuals vary in the values they asign to time spent in various places at various times, and to some extent the total number of trips made through the bottleneck would be affected by the tolls or the congestion conditions, probably in different ways for different users. These complications tend to lessen the sharpness with which critical capacity is determined, but in other ways they may enhance the effectiveness of appropriately graduated tolls and charges improving the efficiency of whatever facilities are constructed.

Expansion of Routes in the Presence of Alternative Routes

One situation that makes appropriately graduated charges even more essential is where a significant part of the traffic has closely competing alternative routes available to it. The classical paradigm of this situation is one where the alternative to the bottleneck route is, for a substantial portion of the traffic, a more circuitous or slower route of ample capacity. In the absence of any charges, the traffic will divide between the two routes so as to equalize total travel costs per vehicle, including travel time and also the queuing time on the bottleneck route. An enlargement of the bottleneck under these conditions will, if it falls short of being able to accommodate all of the traffic, simply result in enough traffic being diverted from the circuitous route to the enlarged bottleneck route to maintain the queue at the former level. The enlargement may thus produce no improvement in travel times at all, at least during periods of peak traffic. In a sense, such a costly enlargement proves worthless precisely because it is free.

In a more general vein, traffic often behaves like population. It has been said that if nothing stops the growth of population but misery and starvation, then the population will grow until it is miserable and starves. Similarly, if the use of private automobiles for access to the cores of large metropolitan areas is so attractive, under uncongested toll-free conditions, relative to other modes, that in effect nothing stops the growth of such traffic but congestion and delay, then such traffic will grow until sufficient congestion and delay are generated to constitute a deterrent, or until the core begins to suffer from gangrene, at which point a cumulative decline may set in that may be difficult to reverse, even with a belated introduction of appropriate toll graduation.

In practice a situation not too far from the classical paradigm often presents itself where attempts are being made to improve access to the core of a metropolitan area. In the absence of pricing, the alternatives may consist of (1) the *status quo*, (2) building an access facility sufficient for the traffic bound for the center, but which will immediately become so clogged with through traffic as to provide only moderate improvement in the speed and convenience of access to the center for a substantial part of the day, and (3) building a huge artery sufficient to take care not only of all local traffic but of any through traffic for which circumferential routes cannot be made sufficiently attractive to divert traffic from the central artery under uncongested conditions, and at the same time trying to bring the circumferentials close enough in and of such high grade as to divert as much of this traffic as possible. This third alternative is likely to prove astronomically expensive as well as disruptive of community amenities; the second alternative may in the end yield a very low return on the investment cost in terms of improvement of traffic conditions. Thus on balance the prospects of a net gain over the *status quo* may be rather dim whatever is done.

The availability of pricing opens up a new alternative of constructing central access facilities scaled to the requirements of traffic actually requiring to go to or from the center, with through traffic during the peak period being fairly thoroughly diverted to the circumferential routes by the charges imposed for the use of the central route. In some instances circumferentials that might not be worth their cost in the face of the difficulty of locating them so as to attract a sufficient volume of traffic in the absence of toll controls will become more worthwhile if pricing is available to help in the optimum distribution of traffic. With pricing such circumferentials may also be more readily located where construction is cheaper and less disruptive of amenities, and possibly also better suited for that part of their traffic that is not in any case tempted to use the central route, without fear that impairing their competitive relation to the central route would lead to undue congestion on the latter.

Variations in the Value of Time

An important but not essential element in the strategic importance of pricing as a factor influencing investment decisions is the existence of variations in the value of time, not only for different persons at the same time, but for the same individual at different times. In the absence of pricing, expansion of capacity must provide indifferently for individuals for whom the improvement will be worth relatively little as well as for those for whom it may be worth a good deal more. Pricing makes it possible to exclude the low-value uses and base the magnitude of the improvement primarily on the uses that are valued sufficiently highly so that they warrant the marginal cost of the final increment to the magnitude of the improvement. The selective effect of pricing on the costs of congestion would be to reduce still further the figures near the bottom of the "Displaced Arrival" cost column of Table 1, enhancing the gains from the earlier increments to capacity and reducing the potential gains from the final increments. Potential improvements in efficiency and savings in construction costs are thus increased significantly over the amounts calculated on the basis of a uniform value for time.

There is, to be sure, likely to be an outcry at this point that pricing discriminates against the poor by forcing them off the congested highways. Actually the number of really poor individuals who are under any strong compulsion to drive cars with any regularity on the congested highways at peak hours appears to be quite negligible. The poorest among those significantly affected by a program of congestion charges are likely to be still somewhat above the poverty line. To the extent that this level of incomes is considered to be in need of a subsidy, there are surely better ways of determining needs than the amount of congested driving done.

A somewhat parallel outcry against the use of appropriately graduated landing fees as a means of controlling congestion at busy airports is even more difficult to justify. It is bad enough when a facility used primarily by the well-to-do is subsidized from tax revenues derived at the margin in large part from taxpayers of lower incomes; at uncongested airports this has at least the virtue of promoting better utilization. But when landing fees geared to congestion costs would substantially improve utilization while costing relatively little to assess and collect, even this excuse is lacking. In the airport case, moreover, those who would be charged the highest tolls, on a per capita basis, would be primarily general aviation planes operated in many cases for company executives and the like or private planes used for recreational and other purposes. Such users would in general be far better able to find acceptable alternatives, such as use of some of the smaller airports, if they are unwilling to pay the appropriate charges, than the patrons of the scheduled airlines. Landing fees reflecting congestion costs at various times could bring about a coordination of use that might well defer for a considerable time the need for resort to costly additional construction, often at less convenient locations. This can come about in part through the diversion of general aviation flights to other airports, through adjustment of airline schedules to reduce the concentration at the peak hours, by the use of larger planes and by scheduling for increased load factors, even without adjustment in the fare structure. If in addition some of the congestion charges can be shifted forward to passengers through the fare structure, some diversion of travel to less congested times and via less congested interchange airports may also aid in alleviating congestion. Such fare adjustments may also be essential if the best allocation of traffic between short-haul air travel and ground transportation, high-speed or otherwise, is to be achieved. A rush to construct additional airports to take care of threatened congestion may prove particularly costly at the present juncture in that improved navigational and flight control methods seem to be on the verge of substantially increasing the capacity of present airports.

Evaluation of Investment in Congestion Relief

Finally, the information provided by a system of congestion control through pricing has an essential role to play in the evaluation of investments designed to afford relief from congestion. In the absence of congestion pricing, very little solid data exists on the value of varying degrees of congestion alleviation, and much of what exists is subject to considerable bias in the direction of overestimating the value of such improvements. For example, a recent study of changes in street use and traffic patterns in central London over the period from 1960 to 1966 came to the conclusion that what was widely touted as a significant improvement in traffic flow was actually no net improvement and possibly a deterioration from the standpoint of origin-to-destination volumes and times.[2] Many of the "improvements" during this period consisted of conversion to one-way traffic and the prohibition of certain turns involving crossing other traffic flows. Although average speeds and flow volumes of vehicles passing given points may have increased, these increases appear to have been used up in traversing more circuitous routes between given origins and destinations. In a similar way, higher speeds and volumes of traffic recorded on turnpikes and expressways often significantly overstate the increase in the transportation service accomplished as a result of their construction in terms of delivery from a specific origin to a specific destination. This is because distances, especially for shorter trips, are often longer via the new routes than via the old, though perhaps not as much as in the London example just cited.

---

[2] J. M. Thomson, "The Value of Traffic Management," *J. of Transp. Econ. and Policy*, Jan., 1968, pp. 3–32.

If, indeed, all routes were subject to appropriate congestion tolls, the level of these tolls would then be a good initial approximation to the value of the congestion relief afforded by investment in increased capacity, at least for small increments. For larger increments one could also then rely, to some extent, on estimates of consumers' surplus under each demand curve separately. But where charges for the use of alternative routes fail to reflect congestion costs at the margin, the problem becomes much more complicated. Not only must allowance be made for the indirect effects on competing routes but consideration must be given to the possibilities for improving efficiency through introduction of appropriate patterns of user charges. In the absence of the information that would be provided by the charging of appropriate tolls, planning of investment in expanded transportation facilities is half blind, and resort is sometimes had to arbitrary rules of thumb, such as that of providing capacity adequate to handle the traffic during the thirtieth heaviest hour of traffic out of the year. The capriciousness of such a rule should be fairly obvious.

Appropriate patterns of congestion tolls are thus essential, not only to the efficient utilization of existing facilities, but to the planning of future facilities.

TRANSPORTATION AND PRICE STABILITY

By George W. Wilson
Indiana University

The question of price stability is generally discussed as part of the theory of inflation. As such it has long been analyzed in terms of various macrostatic and macrodynamic models. Inflation is believed to be either a product of aggregate demand pull or aggregate supply push with the latter stressing labor costs and administered prices by oligopolistic enterprises. When micro considerations are involved we generally get some version of the inter-industry demand shift argument in which prices and wages are relatively inflexible downwards but positively responsive to upward shifts in demand functions or a varient of the "markup" interpretation of inflation. Except for banking, it has not been fashionable to examine one industry as in any sense causally related to the inflationary process. All industries are deemed to react to shifts in cost and demand and to vary in relative efficiency in a somewhat predictable manner once we know their market structures and any pertinent institutional factors associated with them. Thus, as one who has argued (vainly) against the view that transportation is a set of industries inherently different from others, I have some qualms about singling out transport in this connection.

There are three aspects of transportation generally cited that may render its role in inflation somewhat different, qualitatively and quantitatively, from many other industries. First, it is almost as ubiquitous an input item in the productive process as labor—a factor hitherto given great prominence in discussions of inflation. Second, it is subject to a degree of regulation almost as pervasive as those institutions having influence on the money supply—another factor typically stressed in theories of inflation. Third, much of the investment in transport capacity is public and generally less subjected to rational economic criteria and market forces than other areas of capital formation. Partly for these reasons, freight rates have been used to effectuate broad social goals, including such things as regional growth, national defense, antitrust policies, and others. Transport investment, taxation or exemptions therefrom, and direct subsidies have often been used to support overall stabilization policy as well as the goals indicated above. The attempt to utilize transport to effectuate objectives other than efficient provision of this service necessarily leads to misallocation within transportation and between transportation and the economy as a whole. This is a factor conducive to inefficiency over and above whatever wastes may be occasioned by virtue of the "natural" market structure. The mixture of these public and private ingredients may produce a result even more wasteful of resources than either standing alone. It is in this sense that transportation may be viewed in the macrocontext of the general price level.

A Macro Approach to Effects of Transport on the Price Level

Let us postulate an aggregate output function relating the volume of real output (in prices of a given year) to aggregate spending that is linear up to a level of output somewhat below capacity at which point (A) the increment of output with respect to expenditure increments decreases and steadily approaches zero as aggregate spending rises. Up to the point A, the general price level will be constant and equal to the slope of the linear aggregate output function. Beyond *A*, the general price level progressively rises in response to increases in spending. But this causes the aggregate output function itself to shift, or rather pivot, upwards from the origin, which leads to a lower equilibrium level of output at higher prices. This may occasion a further upward shift in aggregate expenditure due to higher money incomes, a possible change in the elasticity of price expectations, and so on. Assuming that autonomous policy changes do not interfere for a sequence of time periods, equilibrium of aggregate output and spending will take place at successively higher price levels and volumes of output that may decrease, remain constant, or increase somewhat, depending upon the magnitudes of the successive shifts in the functions and the shape of the aggregate output function beyond *A*. For any given set of upward shifts in aggregate expenditure for points of intersection beyond *A*, the price increase will be greater the less the elasticity of output with respect to expenditure.

The relevance of any single industry to this process for the economy as a whole depends upon its size and efficiency. The larger the industry and the greater its deviation from an attainable efficiency norm, the greater its influence upon the inflationary process. This I take to be the meaning of the Council of Economic Advisers' statement to the effect that such "industries constitute a further source of inflationary bias in the sense that they offer unexploited opportunities for significant cost reductions" (*Annual Report,* Feb., 1968, p. 99).

Specifically this means in a macrostatic context that the aggregate output function has a greater slope and that the point *A* is further to the left on the output scale. How does transportation in the U.S. influence the slope of the function and the locus of *A*?

Macrostatics

We have long been accustomed to berate regulatory policy because of the chronic failure to relate rates to LRMC which means that the modal split is suboptimal. While there are innumerable problems of measurement involved it is possible that the annual cost involved in such suboptimal modal splits is about $500 million per year[2]—not a large sum relative to GNP but an amount of overspending that evoked the specter of inflation in response to steel price rises some months back. I do not put much faith in this estimate, largely because of the problems involved in assessing the value of the quality differences among the modes for specific traffic and the elusiveness of marginal cost measurement in transportation, about which I have already written too much. Nevertheless, there can be little doubt that users of transportation are paying more than would be the case were freight rates to reflect costs more closely.

There are other sources of static inefficiency within the trnsportation industries which impart an inflationary bias to the economy as a whole.

Technology Gap. In a static sense, this refers to employment of techniques that yield poorer performance than is possible and economically desirable with present knowledge. Here, as well as in the dynamic case, the influence of regulation has unquestionably been perverse. Any regulatory agency must be concerned with questions of equity and, as with the case of the ICC, is mostly devoted to prevention of abuses of monopoly power. These lead to delays in implementing even decisions whose effect is to raise efficiency and to excessive concern for those other firms or regions whose well-being will be adversely affected through a rise in efficiency of an innovating carrier. At any moment of time, then, there exists a backlog of known techniques that would be exploited were regulatory policy less concerned with equity, externalities, and stable interrelationships among regulated enterprises and their customers or what I have referred to as the "grand Benthamite design." Furthermore, to the extent that regulation supplants interfirm rivalry, the incentive to innovate may be blunted.

In general, it is evident that regulation—certainly as now practiced in transportation—is more oriented towards and more effective in the negative role of preventing abuses of economic power than in the positive role of identifying and stimulating or even permitting changes that would enhance efficiency. The legacy is therefore technological backwardness in the sense indicated above.

This does not, however, bear with equal impact upon all the forms of transportation, partly because of unequal regulation and partly because of other forms of public policy designed to stimulate technological change in some modes faster than would otherwise be expected. This in turn implies that the present capital stock devoted to transportation in the U.S. is suboptimal both in amount and allocation among and even within the various modes.

Capital. It is extremely difficult to get a firm estimate of the amount of capital in transportation. However, a tentative figure suggests that about $180 billion is invested in transportation assets both public and private (excluding private automobiles). Net value added by firms providing transport services is roughly $25–27 billion, which implies an average capital-output ratio of roughly 7 to 1. If some account is taken of the noncommercial use and benefits of some portion of the publicly provided rights of way, the ratio would decline somewhat. It is not clear that this ratio is necessarily too high except to the extent that improved performance is possible and that reallocation and elimination of such restrictions as those involving route, back haul, and commodity constraints would enhance efficiency. The ratio is, of course, well above the national average and implies a low average capital efficiency especially when public assets are included.

The differential rates of return among the various modes implies that substantial relative reallocation could enhance overall efficiency in the use of capital. For example, the average ratio of net revenues from transportation operations to net investment in transportation property (including working capital) over the four years ending in 1966, has been 3½ percent for Class I railways, 19.7 percent for Class I intercity motor carriers of freight, 14.2 percent for Class A and B water carriers, 15.2 percent for oil pipelines, and 9.7 percent for domestic trunk air carriers (data from ICC *Annual Report,* 1967 and CAB annual reports).

There are, of course, serious problems in using such data as indicia of misallocation. As far as highways and waterways are concerned, as joint use, multiple-purpose facilities, there is the problem of determining the appropriate share of assets assignable to strictly commercial activities, which the above data involve. In addition, the chronic failure to assess full user charges for use of waterways, highways, and airways, for reasons partly related to their joint use nature but largely due to political pressures, renders the rate of return comparisons less valid as indicators of efficiency. At the same time, of course, such a failure almost guarantees a higher degree of misallocation of resources than needs to be incurred. It is probably fair to conclude that too much capital is currently devoted to transportation and that it is not efficiently allocated among the modes.

Manpower. The differential rates of return noted above may also reflect varying rates of remuneration and/or different relative effectiveness in the utilization of the manpower devoted to transportation. There has been much discussion of the effects of such things as the work rules and seniority provisions pertaining to railway operating employees, the teamsters union, state "full crew" requirements, and so on. In terms of the average level of compensation it is not clear that these have led to excessively high rates of remuneration. In general, labor compensation in rail transportation has increased roughly in proportion to labor value productivity. On the other hand, the work rules suggest that an excessive number of employees are involved and that there is considerable inflexibility in their use. To a much lesser extent, a similar situation prevails in other modes.

Much has been said about the quality of management in transportation, with the railroads and airlines believed to be especially seriously plagued with relative ineptitude in this connection. It is, however, notoriously difficult to evaluate managerial quality particularly in a regulated industry, although it can be done (R. J. Gordon in [6], pp. 60–94).

Direct Regulatory Costs. In addition to the effects of regulation regarding allocation and speed of innovation there are certain direct regulatory costs. Looking only at the ICC, its expenditures were about $27 million prior to the recent budget estimates. There are, in addition, very substantial costs associated with regulatory proceedings, preparation for them, and subsequent litigation and appeal. On very conservative assumptions and using data generated by the Administrative Conference of the U.S. in 1962, such costs related solely to the ICC are probably above recent ICC appropriations. If we include all other agencies involved in transport, an estimate of about $100 million per year does not seem excessive. Not all of these costs would be avoided were economic regulation or its extreme complexity removed, for it is quite likely that considerable litigation under the antitrust laws, especially the Robinson-Patman Act, would ensue. Furthermore, a growing proportion of the regulatory authorities' budgets, including the DOT, appears to be devoted to safety. Thus, far from all of the expenditures now related to economic regulation would result in net savings to society as a whole should deregulation occur. I have little doubt that some savings would ensue but any attempt to pinpoint the amount would be sheer guesswork.

As far as macrostatics are concerned, the situation in U.S. transportation today appears to be that too many resources are devoted to provision of this service. Whether the degree of waste is larger than in other major industries is difficult to judge. What we can say, however, is that transportation in the U.S. under present arrangements raises the slope of the aggregate output function and shifts the locus of A to the left; thus, transportation contributes to inflationary bias in the sense that for any level of total expenditures,the price level is or will be higher than if these inefficiencies in transportation were reduced.

How serious this is, is another matter. The usual figures relating to the significance of transportation regularly overstate the case by comparing total carrier revenues and private transport costs to GNP. In fact, both value added in transportation as a proportion of national income and total employment in the provision of transport services by commercial firms relative to national employment are less than 5 percent. Capital devoted to transportation appears to be roughly 10 percent of total reproducible tangible assets. Thus, while not unimportant nor insignificant, the resources devoted to provision of transportation services by commercial enterprises, including public investments, are not such a large proportion of the labor and capital resources of the nation that enormous gains in total potential output would ensue through improvement in their allocation to and within transport.

For example, if we were to assume that 10 percent of the labor and capital could be removed from the transport industries with zero loss of output and if they were to be considered as net increases in the supply of labor and capital, using a crude version of an aggregate Cobb-Douglas production function with disembodied technical change, the increase of total output would be about six-tenths of one percent. Since these assumptions are extreme, it is probable that total output would rise by much less than this. This suggests that even an extreme pruning of excess resources from transport would not significantly raise total productive capacity. On the other hand, assuming that half of the above resource savings are realizable, along with half of the overspending on transport due to failure of rates to reflect LRMC more accurately and that a quarter of the present administrative costs can be viewed as net savings, total output could rise by about $3 billion at current levels of GNP. Relative to GNP, this may not be much, but it is not an inconsequential amount when viewed against the sums allocated for, say, foreign aid, poverty programs, etc. Furthermore, since the above did not include the labor devoted to provision of transportation by private carriers, which may be as much as that employed by commercial transport firms, the net savings from reduction of inefficiencies may approach $4 billion on reasonable assumptions.

I do not pretend to know whether this is an extremely large sum compared to what might be achievable through possible reductions in inefficiency in other major sectors of the economy (such as defense, education, construction). Some of it may represent a second-best approach and/or creation of positive net externalities not captured in such estimates.

However, if the $3–4 billion is attainable without serious dislocation costs, it is clear that we should not countenance perpetuation of such an amount of needless waste. In the context of dynamic inflationary pressures on the other hand, so long as the relative amount of waste has not changed over time, this would not influence the extent to which transportation has a role in inflation as a process. We must then turn to dynamics.

Macrodynamics

Here I refer to a stretching out over time of the aggregate output function. To test the performance of transport in this connection we need to examine the trends in efficiency. Kendrick's estimates of total factor productivity 1899–1953 place the transport sector ahead of mining, manufacturing, and the total private domestic economy in terms of annual average rates of improvement [3, p. 134]. Barber calculates that from 1947–61 labor productivity in transportation (rail, air, inland waterways and petroleum pipelines) increased from three to five times the national average [1, p. 838]. Mansfield has concluded that the rate of technical change in the railroad industry exceeded the national average for the period 1890–1953 [6, p. 195]. In short, much of the evidence—spotty, inconclusive and conceptually tenuous as it is—suggests a degree of dynamism in the commercial transport sector that would appear on the surface to conflict with the views in the previous section. Yet these findings, even if they could unambiguously be certified as accurate, do not denote optimal nor achievable performance, for it is evident that performance could have been even better, that freight rates have not mirrored the efficiency gains, that serious distortions have been created within the transportation industries which were not inherent in their growth processes, and, thus, that society may still be paying too much for transportation, as argued earlier. The evidence supporting these beliefs includes the almost negligible sum devoted by the transport firms themselves to R and D. The big inventions thus come mostly from the side of equipment suppliers and may therefore apply to only a portion of the problems facing transport firms [1, pp. 846–52]. Suppliers' myopia, as far as the total transport function is concerned, is believed to have caused technological distortions, especially since the firms themselves do so little research. Again, the influence of regulation, as previously noted, has been in the direction of slowing down the adoption of an invention or improved techniques [4]. Although Mansfield did not find the rate of diffusion of an innovation in transport especially slow for air transportation and motor vehicles [5, pp. 37–42] and only slightly slower for rails [6, p. 196], the effect of regulation is to delay innovation in the first place partly because of the reluctance to permit sharp reductions in freight rates that may be feasible with the new technique.

However, what economic regulation takes away, on the one hand, public largess restores in the form of subsidies for technical development, inadequate user charges, tax incentives for additional investment in particular modes, direct subsidies to permit "profitable" operations, and so on. Though on balance and in an aggregative sense the latter may make up or even overcome the barriers to efficiency and growth caused by economic regulation, the failure to coordinate the bits and pieces of policy means that what is taken away is not matched on a one-to-one basis with what is given back. Each mode is not influenced to the same degree by economic or safety regulation, state and local restrictions, or union rules, and there is no attempt to redress the resulting imbalance by various kinds of public aids to transport. Indeed, they just as often aggravate rather than redress these imbalances. Though impossible to document fully, the failure to coordinate or rationalize public policy, at least in its combined restrictive and promotive aspects, doubtless is mainly responsible for the misallocations to and within the transportation industries. The fact of comparatively high overall increases in factor producitivity has not therefore prevented serious distortions.

On the other hand, the belief that too many resources are presently devoted to provision of transportation services in the U.S. may imply some excess capacity which would then tend to depress or hold down prices should aggregate demand increase. In this sense, the presence of excess capacity and resource misallocation might be thought of as fortuitous and one factor constraining bottleneck inflation at least in this sector. This is, however, much like arguing how lucky we were to have enormous unemployment prior to World War II, since it permitted us to conduct the war with little economic sacrifice in terms of reduction of consumer living standards. Furthermore, the excess capital and labor in transport has meant reduced capacity elsewhere and loss of real output. It is unlikely that the dampening effect of this on price increments in transport more than compensates for the opposite effect elsewhere and the lost production. Misallocation of resources over a long period of time can scarcely be justified simply because it may permit costs in one sector of the economy to be held down in the face of rapid changes upward in aggregate demand at some later point in time.

It is probably fair to conclude that in the macrodynamic sense, policies toward transportation have held back the rate at which the capacity of the economy as a whole has grown. How important this has been quantitatively is impossible to judge and, of course, we have until recent years been preoccupied with attempts to make actual output coincide more closely with capacity growth. But once this gap has been closed further increase in aggregate demand will bump against capacity earlier, and hence raise prices sooner, partly because of the economic distortions in the transport sector.

Microdynamics

Any industry suffering from misallocation will have higher production costs than necessary. If its product is an input for other enterprises, this raises the cost functions for all firms purchasing from it. Thus, inefficiency and chronic misallocation in transportation raise cost functions for all commodity producing firms. This is, of course, microstatics and was implicit in the earlier discussion as manifest in the slope of the aggregate output function. We may therefore concentrate on microdynamics.

The evidence on trends of transport prices and the several general price indexes since World War II indicates little relationship especially since 1960. In general, an index of freight rates construed in terms of average revenue per ton-mile shows greater stability and a slower increase than the index of wholesale prices, the consumer price index, and the GNP deflator. Again, there are well-known conceptual problems in these measures, and too much should not be read into them. But the available evidence suggests that freight rate changes have not contributed much if anything to general inflation, however the latter may be defined. This is, of course, independent of the fact that transport costs and freight rates may be excessive in the first place due to the misallocations noted earlier. The postwar rates of change in freight rates are relevant here regardless of initial levels.

Indeed, because of the regulatory process, accommodation to rising costs in transport through general rate increases is subject to rather lengthy delays. This then tends to restrain cost increases in other sectors. Furthermore, the growing competitiveness within transportation, including private carriage, restrains rate increments even for those commodities where the elasticity of transport demand is low.

On the other hand, there is a kind of "announcement effect" of general freight rate increases that may induce many industries, especially oligopolies having some unliquidated profit potential, to use this as an excuse to raise prices earlier and by larger amounts than would otherwise be the case. To the extent that freight rate increases influence the cost functions in oligopoly by similar amounts, this reduces the degree of uncertainty and raises the probability of a noncollusive price increase. In addition, while the regulatory agencies cause substantial lags in granting general rate increases, they also cause delays in granting specific rate decreases, applications for which have become more numerous over time. In this latter sense, an uneconomically high level of rates is maintained.

Finally, freight rates enter into the productive process at several levels and if firms generally engage in cost-plus pricing, a single rate increase may be pyramided substantially by the time it reaches the ultimate consumer. How important this is in the inflationary process is difficult to judge. Much depends upon the extent of vertical integration and the degree of competitiveness within transportation and the industries which purchase transportation. These factors determine both the number of pyramiding opportunities and the extent to which it occurs, especially over a short sequence of time periods.

On balance I am inclined to the view that price increases in transportation contribute to an inflationary process once begun but that this contribution is not much different than in other industries. After all, firms can substitute other inputs for transportation and offset the impact of rising freight rates in various ways (e.g., heavier loading, better packaging, own account transportation).

Conclusion

I doubt that transport can be viewed as in any sense causal or even as a major contributor to inflationary processes. Though beset with problems of serious misallocation for a wide variety of reasons, the overall impact on the capacity of the system is doubtless not large. Though freight rates that differ widely from any computed level of LRMC clearly distort cost structures, locational decisions, and so on, it is hard to believe, considering the ubiquity of a variety of transport facilities and the generally low ratio of freight charges to delivered prices, that such deviations are significant for a large proportion of U.S. industry. Though transport, as presently organized, operated, regulated, subsidized, taxed, and restrained in the U.S., is a long way from achieving either the least cost goal or an optimal growth path, its role in the inflationary process is at best marginal. If society would refrain from treating transport either as a tool to effectuate goals better achieved by more general policy options now available or as an especially critical industry needing the panoply of uncoordinated policies currently associated with it, I am sure that discussions of the role of transport in such things as development, price stability, and so on would become as meaningless as the role of any other industry in such processes. It is the effectiveness of the mechanism of resource allocation to specific industries and the growth in the quality and quantity of resources that are most important in these contexts. Were the U.S. approach to transport less warped by what I have elsewhere referred to as the "grand transportation mystique" (GTM), the present discussion could have been avoided. Price instability is better attacked by monetary and fiscal policy than by using any particular industry to, say, dampen down such a generalized phenomenon as inflationary pressures. It has long been argued, for example, that inflationary pressures cannot persist without an increasing money supply. Society's concern at the industry level should be with efficiency which requires a concurrent improvement in the competitive environment, more rational decision-making processes with respect to public and private investments and more rigorous attempts to close the gap between private and social costs.

Special concern with transportation in these respects is not only warranted but understandable, because economists are generally agreed upon the direction in which changes should be made to enhance efficiency, competition, and rational investments in transportation. Furthermore, most of the restraints are legal or administrative and subject to change if, but only if, some future President is willing to give the indicated changes his full support.

Most of the past approaches to transportation do not have these concerns as main objectives; at least the present policy patchwork does not move society closer to them.

As a final note with reference to poor countries, the role of transportation in inflation may be more direct because a much larger proportion of transport assets is owned by public authorities and deficits in their operation account for high percentages of overall governmental deficits. The lack of a national transport system creates regional bottlenecks and inhibits response by extraregional suppliers which would tend not only to even out regional price differentials but restrain their initial rise. Finally, the policy options of poor countries are more restricted, and thus transport may have to be used to accomplish such objectives as regional growth or income redistribution in the absence of more direct approaches. In this as in so many other contexts, analyses germane to wealthy economies do not have the same relevance to poor ones.

REFERENCES

1. Richard J. Barber, "Technological Change in American Transportation: The Role of Government Action," *Virginia Law Rev.*, June, 1964.
2. Ann F. Friedlaender, "What Course for Transport Regulation?" (manuscript prepared for Conference of Experts at the Brookings Institution, 1967).
3. J. W. Kendrick, *Productivity Trends in the United States* (Princeton Univ. Press for N.B.E.R., 1961).
4. Paul W. MacEvoy and James Sloss, *Regulation of Transport Innovation* (Random House, 1967).
5. Edwin Mansfield, in *The Employment Impact of Technological Change*, Appendix II of the Report of the National Commission on Technology, Automation and Economic Progress (Washington, D.C., 1966).
6. National Bureau of Economic Research, *Transportation Economics* (New York, 1965).

DISCUSSION

Donald V. Harper: Mindful of the fact that to be a critic or discussant of the ideas of others is much easier than being the originator of ideas, I proceed with considerable caution to discuss Professor Farris' paper on "Transportation Regulation and Economic Efficiency."

The first of the major points made by Professor Farris appears to be that economic regulation of transportation is based on the idea that transportation is a monopoly-type activity and that even with the newer forms of transportation the "monopoly pattern" has been imposed, although transportation monopoly has long been a thing of the past.

I can only partially agree with this point, although it is one that has been made by many in the past. It is true that regulation of railroads and oil pipelines had its original justification in monopoly power of the carriers, and that many of the types of regulation imposed on railroads and oil pipelines have also been imposed on motor carriers, water carriers, and airlines. However, it is also true that introduction of regulation of these latter modes was to a large extent the result of a recognized need to control competition within those modes rather than to control monopoly. Hence the strong emphasis on entry control—something which did not come about in railroad regulation until 1920 and still does not exist in the regulation of oil pipelines. Actually, regulation of all modes has been designed to cope with both monopoly and competition, with the emphasis on each problem varying with the mode.

If it is intermodal rather than intramodal regulation that Professor Farris has in mind, then I would point to such things as entry control, the regulation of minimum rates, and the enactment of the National Transportation Policy statement in 1940 as evidence that Congress has at least to some degree recognized that we have a competitive, multimodal transportation system.

A second major point made by Professor Farris is that transportation is oligopolistic in market structure. This is based on the fact that there is a small number of carriers, especially if one considers a given route between two points, and that these carriers are highly interdependent and exhibit some degree of product differentiation.

Whether or not transportation represents oligopoly, of course, depends in part on what one thinks the characteristics of oligopoly are, particularly in connection with the number of firms involved. How few is few?

In any event, I am somewhat unsure as to whether Professor Farris means that transportation is oligopolistic when it is under economic regulation or when it is without such regulation or under both circumstances. My own view is, with reference to the intramodal market structure, that without economic regulation a large part of our transportation system would probably not be oligopolistic but, instead, would be made up of a fairly large number of relatively small carriers. The motor carrier, water carrier, and airline industries would, perhaps, lean more toward oligopoly today than they did in the 1930's, primarily because of the larger capital investment required to enter, but I doubt that they would qualify for the designation of oligopoly without the existence of economic regulation. The principal reason that they appear to take on the characteristics of oligopoly today is the fact that entry is controlled by government. For example, who knows how many airlines there would be without Civil Aeronautics Board control over entry? Perhaps the development of the unregulated third level airlines will provide information along these lines in the near future.

In referring to intermodal relationships, Professor Farris points out that the several modes are all substitutes for one another in a "substantially relevant range" and, therefore, illustrate a situation of oligopoly. It seems to me that, although the modes are sometimes substitutes for one another, it does not appear that this is a good example of oligopoly. The economic characteristics of each mode and the services they offer are quite dissimilar and, therefore, I prefer to think of the five modes as separate industries with some cross-elasticity of demand between them. To refer to them collectively as an "oligopoly" is roughly equivalent to saying that the steel, aluminum, and lumber industries together represent one colossal oligopoly because some of their products are substitutes for one another. The oil, natural gas, and coal industries represent another such collective example. I don't believe that oligopoly theory is intended to apply to such situations, but rather to relationships between firms in the same industry.

A third point made by Professor Farris is that economists have continued to plead for more competition in transportation, even though transportation is an oligopoly, and have failed to apply oligopoly theory to transportation.

But suppose that transportation is an oligopoly and that economists recognized it, could they not still plead for more competition? Such a plea would not mean that economists are advocating extremely large numbers of small carriers with no product differentiation and with complete freedom to enter and exit. It would mean, and has meant, instead, I believe, that economists and many other people are asking for more freedom to compete in transportation by eliminating some of the regulatory shackles that now restrain the carriers. Does not the oligopoly market structure model allow for freedom to compete? I think it does.

Indeed, the fact that oligopoly allows for freedom to maneuver and that it assumes independence of decision making on the part of each seller make it difficult to apply oligopoly theory to regulated transportation. Competition, particularly price competition, and independent decision making are hampered when economic regulation exists. Therefore, it appears that the oligopoly that is in part created by regulation is not suitable for the application of the theory of oligopoly, at least in the conventional sense.

A small example of this problem is found in the rate bureau proceedings mentioned by Professor Farris. Would it be possible to apply oligopoly theory to pricing in a rate bureau when legalized collusion is the rule among the few competitors and the government stands ready to "correct" any "wrong" decision made, if necessary?

Actually, the plea for more freedom to compete has little to do with whether or not we call transportation oligopoly or monopolistic competition or whatever. It merely means that many people think it would be desirable to permit more freedom to adjust prices and services to the needs of the market with less interference on the part of government. There is the belief held by many that economic regulation, both intramodal and intermodal, now results in a great deal of economic inefficiency and that more freedom to compete would change this situation for the better. Restrictions on the freedom to compete that are found in entry control, rate regulation, and other aspects of regulation can be a cause of economic inefficiency in our transportation system, since they can have the effect of slowing down decision making, making it difficult for carriers to adapt quickly to changes in their environment, misallocating traffic among modes, forcing empty back hauls in some situations, providing protection to the inefficient carrier, making rates higher than they would be without regulation, making single-carrier service over long hauls difficult, discouraging initiative, preserving excess capacity, and so on. The degree of inefficiency caused by regulation is, of course, not equally distributed among the several modes. And, in addition to the problem of inefficiency, it also appears that the goals of regulation have been lost sight of in the tangle of laws and commission and court decisions that have accumulated over the past eighty years.

Economic regulation, regardless of its effects on economic efficiency, is here to stay for a long time. What is needed today is discussion and action relative to what the objectives of regulation should be now and in the future, and how the patchwork of economic regulation is to be revised and brought up to date so that it accomplishes those objectives with as little economic inefficiency as possible, or better still, while promoting greater economic efficiency than we now have.

John L. Hazard: Professor Wilson appears to be going against his own principles in singling out transportation for separate consideration in the general problem of price inflation. If this is the case, we are all beneficiaries of his temporary lapse of principle. For it does set the transportation problem in a new and broader context. I hope to persuade him to continue the lapse long enough to update and enrich our deliberations.

Has transportation added to general price inflation? Not very much, according to Wilson. But we should not be sanguine about this for "carriers could be doing much better." And they could do better if certain changes generally agreed by economists were instituted to enhance efficiency, improve competition, and rationalize investments in transportation. "Furthermore, most of the restraints are legal or administrative and subject to change if, but only if, some future President [written preelection] is willing to give these indicated changes his full support." But we should, in any case, avoid the "grand Benthamite design" which he finds justifiable only in poor economies. I suspect that the author employs the term "poor" instead of "developing" nations out of some genuine concern for the plight of their design rather than the poverty of their circumstances.

This is a new and substantive paper in which I find much that is agreeable. Therefore, I shall confine my remarks to the limited areas of disagreement and constructive suggestion.

In holding to the belief that transportation is not inherently different from any other industry with respect to its influence on price inflation, Wilson appears to be questioning the rationale for his own paper. And this preoccupation is, in my view, quite unnecessary, because transportation is inherently different—at least over the macrodynamic period suggested in his paper.

While, in the macrostatic model, the diseconomies of transportation may be conceived as being passed on directly to other industries, thus producing an upward bias to the aggregate production function and consequent higher prices, I am very doubtful that this same phenomenon merely repeats itself over the longer-run adjustment period. Instead, there is a basic disruption in regional specialization and exchange over distance, producing fundamental changes in industrial responses and structure. Industries facing persistent inefficiencies in public transportation will substitute their own carriage, localize resources, and/or circumscribe markets. This process of substitution continues until costs at the existing location exceed anticipated costs at a new location and then the industry shifts or decentralizes. As industry continues to substitute other factors and shift location, transportation—and particularly publicly available transportation—becomes a lesser part of the whole market basket. With substitution the costs of all other factors go up with corresponding upward shifts in price levels, which are passed on with some pyramiding. With shifts in location and regional decentralization, some of the economies of scale possible with a mass market are lost, including those of the railroads. In short, the diseconomies of transportation can be viewed as a tariff imposed between the states and objectionable for the same reasons. Increases in tariff or transport revenues will never exceed the loss of the gains from specialization and trade. Since I can think of no other industry that experiences the range substitutions or occupies quite this strategic position and leverage in the total economic process, I would argue that transportation is different and deserves separate treatment even in examining inflation.

There is also some evidence to suggest that such an autonomous transfer of resources out of public transportation may already be in motion, leaving little excess of resources to be reallocated, as suggested by the author, but rather obsolete plant to be written off. There has been an appreciable shift from public to private carriage, freight revenues have not kept pace with the growth of national incomes, and for the past decade the growth of intercity freight ton-mileage has failed to parallel gains in industrial production or real gross product. Whether this trend is in response to regional population shifts, technological changes in unit economy of scale, or declines in transportation efficiency needs further investigation before developing dynamic models and examining the influence of transportation on price or production trends. If such an investigation were conducted, I suspect that it might reveal that the influence of transportation on price inflation over time has been occasioned not so much by increasing rates as by cuts in the service offering.

As a matter of fact, post-World War II indices of freight rates by railroad reveal a decade in which freight rates preceded other price increases (reaching 121 in 1958 with wholesale prices only 115.7 of 1950 levels). This has been followed by a decade in which freight rates have lagged substantially behind other price increases (dropping to 105 in 1967 with wholesale price indices reaching 122.2 of 1950 levels). One might conclude that railroad pricing has actually restrained inflation were it not for appreciable shifts in the service offering. But during this interval, railroads have gone out of the LCL business, turned pick-up and delivery services over to forwarders, instituted container and trailer services, jumbo cars, unit trains, and rent-a-train services. They have, in short, shifted to a restricted service offering, primarily, to hauling carloads in trains with little else. The more discrete services have not been dispensed with altogether, but rather passed on to other carriers and back to users. Carriers holding out a more complete service package (primarily highway, air, and forwarders) have assumed part of these functions and the accompanying traffic. And, insofar as we can tell from the modest evidence available, their gains in productivity have enabled some restraint in rates. So the carriers on the rate record would appear to have had a restraining influence on general price inflation. The only major exception is public passenger transit, where rates have greatly exceeded the consumer price indices and even here the burden has been assumed primarily by the private automobile whose costs of ownership, repair, and fuel for the past ten years have tended to follow rather than precede the consumer price index. So it is only the traffic and services passed back to the user and small shipments that escape our view entirely. This submerged portion of the total cost, often referred to by economists as "nontransportation costs," is, in my opinion, the major part of the iceberg.

How much of the iceberg is hidden from view is difficult to measure. Studies by the ATA and Heskett estimate that the total freight transportation expenditures may run in excess of $70 billion a year or 10 percent of GNP. Half of these expenditures are made by nonreporting private and exempt carriers. If this is the case, then, Professor Wilson is apparently viewing less than half of the freight picture when he alludes to the $27 billion in values added by firms providing transport services or the 5 percent of GNP provided by commercial firms. His analysis of potential savings could be correct only if he assumes perfect allocation in the private sector. The error becomes more serious when he includes public capital committed for highway, air, and waterway systems in his total $180 billion invested in transportation assets in the divisor and only the values added by commercial carriers as the dividend ($25–$27 billion) to derive a capital-output ratio of 7–1. If values added by noncommercial carriers (private and exempt) were included, the capital output ratio may have been 2.5–1 or 3–1, more akin to railroads. If automobiles had been added to the capital and values added by automotive services to the dividend, the capital output ratio may have been even lower. Would the inclusion of these omissions have caused him to revise his ideas about the excesses of resources employed in transportation and to modify his analysis accordingly?

Then there is the part of the iceberg that we have all tended to overlook by entitling nontransportation costs. The Department of Commerce estimates inventory carrying costs to be around $30 billion a year, or 4 percent of GNP, and this excludes related costs of handling and distribution. How much of this transport related cost has been passed back from carriers to users? What are the trade-offs between logistics-transportation-distribution costs at this level? How much of the initiative in transportation has shifted with the function from the carrier to the user? And how, if at all, does the shift affect industrial costs and pricing? It seems to me that we shall have to contemplate these questions before we can say much about the dynamics of the transportation function, including its influence on pricing at either the macro or micro level.

Finally, Mr. Wilson on prior occasions has mourned the decline of ideological differences in the field of transportation. I would be remiss if I did not remind him that the victim is battered almost beyond recognition but still sufficiently alive to resist burial eulogies. Mr. Wilson and, indeed, many economists attribute most of inefficiency and misallocation in transportation today to warped and uncoordinated governmental policies, subsidies, and regulations. If these instrumentalities of "the grand Benthamite design" more appropriate to an underdeveloped country were removed, a competitive model would evolve in transportation to make the harmonious allocations necessary in an advanced industrial economy. While I am favorable to the direction of movement, I can foresee many problems of transition. If, for instance, MC rates and rules were implemented, the ICC might pare its expenses by a quarter, but I would visualize at least a concurrent increase in cost accounting and court costs. The regulatory agencies might mitigate their parental role only to find the antitrust docket filled with new cases as large industries absorb carriers for use as a competitive instrument. The government might use harder criteria for evaluating public investments in transportation and find itself confronted with increasingly serious unmet social needs. Some of the problems of the ghetto, urban redevelopment, poverty areas, resource development, economic growth, and environmental improvement are amenable to other forms of support, but some are more directly and economically accommodated by alterations in the transport system. While private capital in pursuit of profits can be relied upon to make many of the major changes, there will remain many areas of congestion, strikes, rapid technological advance and uncertainty in which the government has the mandate, resources, and powers to initiate change. That government should preserve its multiple initiatives and work far more with and through industries, state and local governments, universities and research agencies, I have no doubt. But the public interest in transportation will not disappear, in my estimate, but merely change its form and locus. To assert otherwise may be an attempt to confine the problem to the present limits of the discipline. But I suspect this approach may be a contemporary manifestation of transportation economists' rebellion against the establishment which may turn out to be but a Freudian death wish. There is no shortage of systems engineers, public planners, or social scientists to fill the gap.

JOSEPH V. YANCE: My assignment is to comment on Professor Vickrey's paper. I would first like to mention some of the other work on congestion theory and investment related to his paper. Mohring, in a 1962 book on highway benefits, and, later, Strotz and Vickrey, analyze the case of highway investment for a steady traffic demand, using the usual congestion model where average travel time is an increasing function of the hourly volume. One of the interesting results of the analysis is that if there are constant returns to scale in highway investment, and if you have congestion tolls calculated in the usual way (equal to the difference between marginal social cost and average cost), then for an optimum solution, toll revenues will just cover the road cost.

Professor Vickrey's present paper considers the matter of a time-varying pattern of demand. To simplify the analysis, he assumes the facility is a bottleneck; that is, one with a sharply defined capacity.

I wish to comment on two things in the model: first, the way he obtains an actual commuting pattern and, second, the implications of the model for investment.

I was interested in the derivation of the commuting pattern since it is based on an optimization rationale for individual commuters, and would like to spend a minute going over the derivation. The total benefit to the individual commuter of going to work in the morning is assumed to be a linear function of the time spent at home, the time spent in the line waiting to get through the bottleneck, and time either spent at the office early for work or time late at the office. For example, for someone arriving at the office early, he gets a utility function which, when we differentiate with respect to arrival time, yields the marginal condition for some-

TABLE 1

Investment Analysis (Road Cost $600/Lane/Day)

v	Road Cost	Displ. Arriv.	Total Real Cost	Congestion Tolls	Total Real Cost (with no toll)
50	$1,000	2,016	3,016	3,456	6,472
60	1,200	1,440	2,640	2,880	5,520
70	1,400	1,029	2,429	2,469	4,898
80	1,600	720	2,320	2,160	4,480
90	1,800	480	2,280	1,920	4,200
92.952	1,859.0	419	2,278	1,859.0	4,137
100	2,000	288	2,288	1,728	4,016
110	2,200	131	2,331	1,571	3,902
120	2,400	0	2,400	0	2,400

one arriving early, $dq/dt=1/2$, where $q(t)$ is the time spent waiting in line. That is, for someone arriving early, the time in line would be growing. Similarly, for persons arriving late, we get the second condition, $dq/dt=-1$. This tells the conditions that hold for the individual commuters. Vickrey shows that if the waiting time at the bottleneck increases and falls according to these equations, none of the commuters can improve his schedule, so these equations are a solution to the commuting time pattern. The people due at work earliest will go through the line first and arrive before they are due, and, conversely, for those who are due to start work later in the commuting period.

I am still surprised that the optimization conditions for individuals are sufficient to obtain the behavior for the whole mass of commuters.

I think this approach of determining the demand for roadway use as a function of delay time is potentially very useful—set up the desired time pattern, and the optimization rationale for the individuals, and use these to obtain the actual commuting pattern. I used a similar linear optimization model applied to the question of what flight people will choose, based on a desired arrival time and costs similar to those Vickrey uses—cost of arriving earlier or later than desired and cost of flight time—and obtained some results in empirical application.

My second comment has to do with the optimum investment in the model. First of all, let me try to present the investment analysis somewhat more systematically. If no toll is charged, the real costs are road investment cost, displaced arrival time, and congestion cost. This is indicated in the last column of the Table 1 for a road cost of $600/lane/day. In this no-toll case, the minimum total cost is obtained by having a road with a capacity large enough to completely eliminate congestion, 120 cars/minute.

However, if a congestion toll is charged, then time in line can be completely eliminated in his model by encouraging commuters to arrive at the proper time. In this case, the real costs are the costs for the road and the cost of displaced arrival time only (column 3). To minimize this total cost, we build a roadway with capacity of 93 cars/minute. There is a modest saving in total real cost over the no-toll case.

An interesting result is that at this optimum road size, the congestion tolls just pay for the road cost. Thus at the optimum capacity of 93 cars/minute (92.952), the toll revenue of $1,859 just equals the road cost.

This result holds in general for Professor Vickrey's model. Let us assume that road costs are proportional to capacity ($=cv$). With tolls, the total real costs are the sum of the displacement costs, as given by equation (10), and the road cost,

$$\text{Total cost} = N^2(r/2)(1-r)(w_f - w_p) \cdot ((1/v) - (1/v_m)) + cv. \tag{1}$$

After differentiating with respect to v and setting the derivative equal to zero, the optimum road size v_0 is given by the equation:

$$\begin{aligned} v_0^2 &= (1/c)N^2(r/2)(1-r)(w_j - w_p) \\ &= (1/c)N^2(r/2)(w_h - w_p). \end{aligned} \tag{2}$$

Congestion toll revenues can be expressed in terms of the peak delay, $q_p(1/2)(N)(w_h - w_q)$, or from equation (6),

$$\text{Toll revenue} = (N^2/v_0)(r/2)(w_h - w_p). \tag{3}$$

We wish to show that toll revenue $= cv_0$. This can be done most easily by multiplying each side by v_0, to see that

$$\begin{aligned} &(\text{Toll revenue})(v_0) \\ &\qquad = cv_0^2 = N^2(r/2)(w_h - w_p). \end{aligned} \tag{4}$$

This result—that for the minimum real cost solution, toll revenues equal road costs—is similar to that obtained by Mohring, Strotz and Vickrey for the case of a constant level of demand.

Let me sum up. The assumptions regarding demand are quite restrictive (demand is given and the only question is when someone will go to work). Also, the assumptions regarding the technology or production function of the congested facility are also restrictive, since a bottleneck is assumed. Nevertheless, some progress has been made. The theory of pricing and investment under variable demand has been worked out for utility companies (Steiner, Williamson, etc.), but not for the case of congested facilities, and it is in this connection that Vickrey's paper breaks new ground.

LIBRARY UNIVERSITY

MONETARY THEORY

THE LAG IN MONETARY POLICY AS IMPLIED BY THE TIME PATTERN OF MONETARY EFFECTS ON INTEREST RATES

By Phillip Cagan *and* Arthur Gandolfi
Columbia University

I

One of the controversial and pressing questions of monetary policy is the nature and length of the lag between the execution of monetary measures and the effect on output and employment. While monetary theory suggests many possible reasons for such a lag, there is no general agreement on its length. The work of Friedman and Schwartz [6] [7], who relate cycles in monetary growth to cyclical peaks and troughs in business activity, suggests a lag of from two to six quarters or more. Others claim it is shorter. There are problems, however, in inferring lags from the timing of turning points alone, in part because of possible feedback of income changes on the money supply. It is therefore desirable to bring in additional evidence.

The evidence we examine on the length of the lag is provided by the time pattern of monetary effects on interest rates. There may be feedback from interest rates to the money supply, but the feedback problem appears to be much less serious here. For present purposes, monetary policy is fully represented by the behavior of the money stock. We ignore any lags between policy actions and the effect on the money stock.

A long-standing proposition of monetary theory is that an increase in the money stock reduces interest rates. To some extent this effect is a by-product of an expansion of the banking system. In an open market purchase of securities, the Federal Reserve increases the money stock and bids up the price of government bonds. In monetary theory that effect is not simply an institutional characteristic of our monetary system but reflects the functional relation of money demand to interest rates. Since a variety of other assets can substitute for money, the public attempts to purchase other assets to substitute for the newly issued money. The resulting general increase in asset prices and corresponding reduction in interest rates induces the public to hold the larger balances.

This effect of a monetary change on interest rates is not permanent. The larger balances and lower interest rates induce greater spending, particularly on consumer and producer capital goods. The ensuing increase in expenditures raises income and works to reverse the initial effect of the monetary increase on interest rates. If there is full employment, prices rise and reduce the real value of money balances, which tends to raise interest rates in the same way that the initial increase in balances reduced rates. If the expansion starts from less than full employment, part of the increase in income will be in real terms. A given rise in nominal or in real income will both tend to raise interest rates, but not necessarily by the same amount. A change in real income may shift the aggregate saving and investment curves by different amounts, and so affect interest rates, while a change in income due solely to price changes is less likely to affect interest rates, at least not permanently. The equilibrium level of interest rates eventually reached will therefore depend on how the investment and saving functions shift in response to a change in real income. These complications aside, it is widely accepted that a delayed response of spending to interest rate changes is the main reason for a lag in monetary policy. Such a lag implies a particular pattern to the accompanying changes in rates: with a change in the money stock, short-term interest rates supposedly adjust rapidly, which initially accommodates the economy to the change in money balances. The ratio of money to nominal income rises and produces a decline in interest rates. The lower interest rates induce an increase in expenditures and income which continues until interest rates and the ratio of money to income are returned to their equilibrium levels. The change in direction of interest rate movements following a monetary disturbance reflects, and therefore provides a measure of, the lag in monetary effects on income.

The advantage of measuring the lag by the interest rate pattern is that we do not presuppose any one view of the monetary system. It is justified within a wide variety of theoretical models. The recent article by Donald Tucker [10], which derives such a pattern of response of interest rates theoretically on the assumption of a lag in investment expenditures, is one possibility. One of us has shown that such a pattern also results from lags in adjustment between desired and actual money balances [3]. Laidler [8] obtains similar results using permanent income in the money demand function. The FRB-MIT model shows such a pattern in simulations [9] because of lags built into many of the equations. The only crucial assumption is the widely accepted one that the economy adjusts to monetary changes through a combination of changes in interest rates and income. No assumptions about elasticities are necessary. This pattern will overstate the lag in monetary effects on income only if income changes have a lagged effect on interest rates. We shall assume here that the increased expenditures which account for monetary effects on income are financed in substantial part by borrowing and so at the same time (or even earlier) have corresponding effects on interest rates.

II

Interest rates and income do, of course, reflect a variety of economic developments that are not all of a monetary nature. The statistical association can be affected by the particular form of the equation and the variables included. In an earlier study [1] one of us found a negative association between concurrent changes in interest rates and the monetary growth rate, where money was defined as currency outside banks plus demand and time deposits of commercial banks. To the extent that interest rates influence the supply of money, the association would have to be viewed as a two-way relation. That study examined the data for indications of the reverse effects of interest rates on the money supply and found them to be of minor significance for the period since 1907 as a whole. Interest rates would if anything have a positive effect on monetary growth. The observed negative relation can be attributed to independent effects of monetary growth on interest rates.

Some later work with these data [2], in which lagged values of the monetary variable were included, indicated that the concurrent negative effect of money on interest was gradually offset and subsequently became a positive effect. William Gibson carried these results further [4] and found that a complete offset took three to twelve months or so. Gibson's estimate of the lag varied among the various forms of the function he fit. His data covered the period since World War II only. We have extended these previous findings by covering a longer period of time and by using a statistical procedure to constrain the lag pattern.

One of the questions that arises in this work is whether to use the level or the rate of change of the variables. Our method is to relate the level of interest rates to the growth rate of the money stock (defined to include time deposits), using first differences of these variables in the regression equation. In effect, therefore, we relate changes in interest rates to changes in the monetary growth rate. We see considerable justification for this form of the regression equation. Whether a monetary change affects interest rates depends upon the speed of adjustment of income. If monetary changes are fully anticipated, one would expect people to take account of the effects ahead of time, so that prices would adjust without a lag. If so, there need be no effect on the real rate of interest. It is hard to determine to what extent monetary changes at any particular time are anticipated, but presumably a steady growth rate will sooner or later come to be reflected in a corresponding rise in prices (allowing for the growth rate of real income). Consequently, changes in the monetary growth rate will tend to produce, every time they occur, a response in interest rates, but the same will not be true of changes in the money stock that reflect a constant rate of growth. To be sure, the regression equation assumes that the pattern of that response is always the same, whereas a different pattern at various times is conceivable. That is a much weaker assumption, however, than is implied by the alternative that every monetary change is supposed to affect interest rates without regard to the prior behavior of the money stock.

What exact patterns should we therefore expect to find? Suppose that interest rates are in equilibrium with some rate of monetary growth and a corresponding rate of growth of nominal income, and that monetary growth then jumps to a higher rate, due to a change in monetary policy or for any other reason, and remains there. The increasing money balances are accompanied by the bidding up of security prices, and interest rates will begin to decline. Eventually the rate of income growth will rise, as the lower interest rates stimulate investment expenditures and the larger balances lead generally to greater spending. The initial decline in interest rates will taper off and then be reversed. There are two reasons for expecting the reversal to carry interest rates beyond their initial level. First, nominal income must grow faster than money balances at some point, because the ratio of money to income initially increases as interest rates fall and the ratio will be reduced when interest rates subsequently rise. The adjustment path may conceivably involve some overshooting, in which case interest rates approach the new equilibrium level with damped oscillations.

Second, a higher rate of monetary growth will sooner or later be reflected in a higher rate of price change, which in turn will come to be anticipated to some extent by the public. One of the adjustments to anticipated price increases is a rise in nominal interest rates to compensate for the depreciation in real value of fixed-dollar assets. Eventually the rise in nominal interest rates should equal the rise in the monetary growth rate; but, given the difficulty of anticipating price changes, the full adjustment could well take a very long time. We might expect to measure just the first part of such an adjustment; nevertheless, the lag pattern would, after its initial decline, sooner or later cross above the zero line.

Let us now turn to the evidence. Our regressions are of the form

$$\Delta i_t = \alpha + \beta_0 \Delta \dot{m}_t + \beta_1 \Delta \dot{m}_{t-1} + \beta_2 \Delta \dot{m}_{t-2} + \cdots + \beta_n \Delta \dot{m}_{t-n},$$

where i is the interest rate, $\dot{m}_t$ the monetary growth rate in period t, α and β are regression coefficients, and the operator Δ indicates the first differences of the variables. The equation relates the change in the interest rate during period t to monetary growth rates in t and previous periods back to $t-n$. The coefficients β_k represent the effect on the interest rate in period t of a change in monetary growth which occurred k periods earlier.

Our chief concern is the length of time in which an initial fall in the interest rate due to an increase in monetary growth is subsequently offset by an induced rise in the rate of income growth, and conversely for decreases in monetary growth. Since the regressions are in first-difference form, the sum of the coefficients indicates the cumulative effect on the level of interest rates of a change in monetary growth over a given period. For example, a permanent change in the monetary growth rate affects the interest rate by β_o in the initial period, by β_1 in the next period, and so on. The sum of the coefficients gives the total effect of an increase in the monetary growth rate which starts in period $t-n$ and remains at the same rate for the subsequent n periods.

The top of Chart I presents the cumulative lag pattern of the commercial paper rate for the period 1910–65 and for two subperiods. These patterns are based on regressions of monthly data. Monthly data for the money stock begin in 1907. The first observation of the independent variable included in the regressions is the change in the monetary growth rate from June to July, 1907, and the first observation included of the interest rate is the change for the thirty-ninth month later.

The chart should be interpreted as follows: Suppose the monetary growth rate increases by 1 percentage point. At first the interest rate declines and later begins to rise. For the regression of the full period 1910–65, the initial decline continues for seven months and reaches 2.6 basis points. The ensuing rise in the interest rate takes it past its starting level by the fifteenth month. We cannot tell, from these patterns alone, when income first responds to the monetary change. It may respond immediately, though at first not by much. Presumably, however, the induced rise in income accounts for the upturn in the rate which occurs after six months. The rate recovers over half of its initial decline by the thirteenth month, all by the fifteenth month as noted, and reaches a peak of 3.2 basis points above its initial level in the twenty-fifth month. For a decrease in the growth rate of money, the appropriate pattern is simply the inverse of those shown.

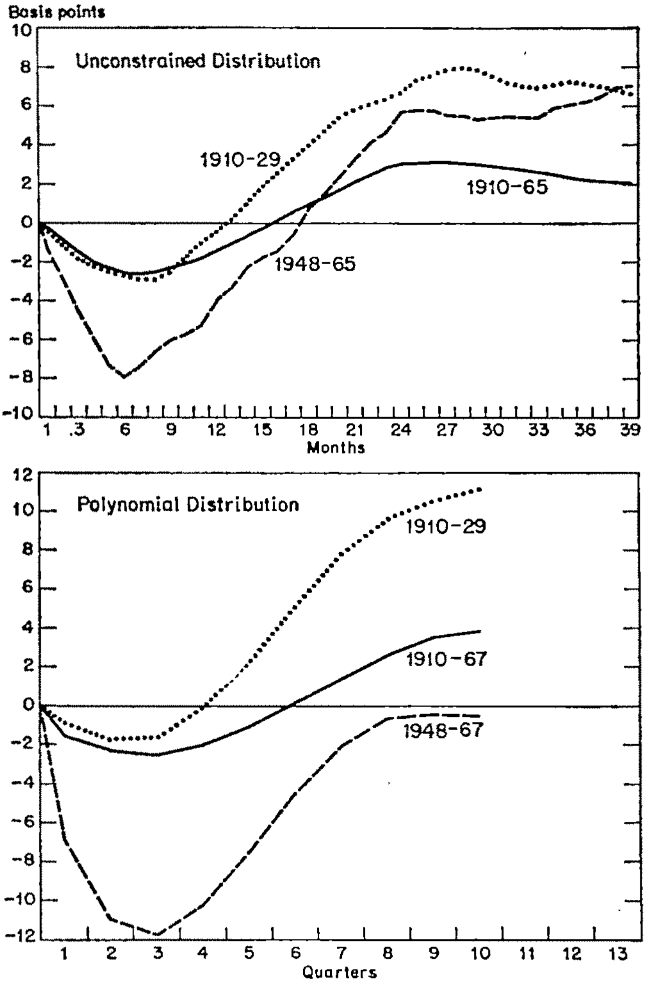
The other plots in the top panel of Chart I show how the pattern differs between the earlier and later subperiods. There is a smaller initial decline in the earlier than in the later period, and a more rapid rise thereafter.[1] All three curves reveal a strong positive effect still remaining after thirty-nine months.

Nineteen of the first 23 coefficients are significant at the .05 level in the full-period regression (see Appendix table). None of the last 16 are. The multiple R is .37 and the adjusted R-square is .08. A low correlation coefficient here should not be surprising, since the regression measures only one of many influences on interest rates and is in first differences.

---

[1] An F test of differences between the regression coefficients for the 1910–42 and 1942–65 periods was not significant, however.

CHART I
THE LAG DISTRIBUTION OF MONETARY EFFECTS
ON COMMERCIAL PAPER RATE



Note: Units are basis points per one percentage point change in annual rate of monetary growth (100 basis points = 1 percentage point of interest rate).

With so many lagged terms, the problem of collinearity among the independent variables is a matter of concern. Here, however, only 109 of the 1481 elements of the correlation matrix have values over 0.1, and the vast majority of these are under 0.2. That is one advantage of running the regression in first-difference form. The shape of the lag pattern is nevertheless affected to some degree by the number of terms included. Experimentation with various lengths suggests that the point at which the pattern rises above zero is little affected by extending the lag beyond twenty months. The results shown in the top panel of Chart I with 39 terms should therefore be long enough to give reliable estimates at least of the first part of the pattern.

We also experimented with regressions

which omitted those years in which unusually large changes in monetary growth occurred; these were the years 1918–19, 1929–32, and 1942–48. In no case did the deletions make a significant difference in the pattern of the regression coefficients.

It is important to emphasize that these lag patterns were not imposed upon the data in any way. The least squares procedure minimizes the variance of the residual terms but does not constrain the regression coefficients. Yet the patterns generally conform to our theoretical conceptions outlined above. The patterns at the top of Chart I appear to approximate a polynomial curve of fairly low degree. As a further check on the position of the change in direction and zero crossover of the lag, we can make use of the presumption that the lag distribution is smooth and constrain it to follow a polynomial curve of low degree. The results of using a fifth-degree polynomial are presented in the bottom panel of Chart I, where the monthly data were converted to quarterly averages because of size limitations of the computer but cover two additional years to 1967. The program we have used was written by Robert Hall,[3] based on the Almon method.

These results are quite similar to the unconstrained lags. Experiments not shown with polynomials of much higher degree and with many more lag terms gave very similar patterns. Table 1 summarizes the fits in terms of the number of quarters taken by the lag to change direction and to cross the zero line, a point at which the initial decline in the interest rate has been completely reversed. By these results, the lag in monetary policy has generally lasted up to 6 or 7 quarters and, for 1948–67, over 9 quarters, not out of line with Friedman and Schwartz's results based on turning points. Taking the bottom of the decline in the interest rate as the beginning of the main response of income, we find the initial response to occur after about 2 quarters. Our estimates of the lag exceed those reported by Gibson [4] for the post-World War II period.

The polynomial lags tend to heighten the differences between the subperiods in the magnitude of the decline and subsequent rise in the interest rate. In a comparison of the 1910–29 and 1948–67 subperiods, the later period displays much larger movements and, apparently for this reason, crosses the zero line much later. The polynomial lag for 1948–67 is also considerably longer than the unconstrained lag covering roughly the same period, for reasons which we have not yet determined. In any event, the implication seems to be that monetary changes have impinged much more strongly upon financial markets since World War II than formerly and, perhaps as a consequence, have involved larger changes in portfolios which worked to produce a long lag in the effect on income growth. Whether this result is simply a statistical artifact of the data requires further analysis.

TABLE 1

THE LAG IN MONETARY EFFECTS ON THE COMMERCIAL PAPER RATE: ESTIMATES OF THE TIME REQUIRED FOR THE LAG TO CHANGE DIRECTION AND TO CROSS THE ZERO LINE

Period Covered	Polynomial Lag Pattern		Time Required for Lag to:	
	Degree	Length	Change Direction	Cross the Zero Line
1910–67	10	20 Q	3 Q	6 Q
	5	15	3	7
	4	8	3	7
	5	10	3	6
1910–29	5	10	2	5
1910–41	5	10	3	6
1948–67	5	10	3	9+

---

[3] In collaboration with John Brode and Frank Ripley. On the lag method, see [5].

IV

It seems desirable to distinguish between the effects on interest rates of a rise in income and an increase in anticipated price changes. The latter affects only nominal, not real, rates of interest, though both still reflect monetary effects on aggregate demand. To the extent that anticipations depend upon the past behavior of commodity prices, the effect on nominal interest rates will also be related to earlier changes in monetary growth which produce changes in prices. The tail-end rise of the lag patterns as noted earlier may well reflect this delayed adjustment of nominal rates to past effects of money on prices and thence on anticipations. We can test for such an effect by adding to the equations a lagged pattern of price changes to represent anticipations.

Our examination of this question is not yet completed. Preliminary results suggest that inclusion of past price changes affects the lag pattern for money in the expected direction. That is, the inclusion removes the tail-end rise in the pattern above zero, suggesting that the monetary effect on interest rates, apart from anticipations of long-run price changes, does not involve overshooting and is not permanent. The inclusion does not appear to affect the first part of the lag pattern.

V

In summary, the lag in the dependence of interest rates on monetary growth is indicative of the lagged effect on expenditures and income. This indirect measure of the income lag of monetary policy suggests that the main effects occur over a period beginning six months later and lasting over two years. This is consistent with other evidence pointing to a long distributed lag. Our results pertain to the average lag and do not indicate how variable it is over time. The problems such a lag pose for an effective flexible monetary policy are obvious, and raise serious questions about its use for offsetting short-run fluctuations in income. To have confidence in the appropriateness of particular policy measures, we need to know the pattern of income responses which they will produce and what would happen with no change in policy.

REFERENCES

1. P. CAGAN, "Changes in the Cyclical Behavior of Interest Rates," *Rev. of Econ. and Statis.*, Aug., 1966, reprinted as Occasional Paper 100, N.B.E.R.
2. ———, "The Channels of Monetary Effects on Interest Rates" (N.B.E.R., 1966, mimeographed).
3. ———, "The Nonneutrality of Money in the Long Run: A Discussion of the Critical Assumptions," Conference on Money and Economic Growth, Brown Univ., June, 1968 (forthcoming in *J. of Money, Credit, and Banking*).
4. W. GIBSON, "Effects of Money on Interest Rates," Staff Economic Studies (Board of Governors of the Federal Reserve System, Jan., 1968).
5. R. E. HALL, "Working Paper No. 7" (Dept. of Econ., Massachusetts Inst. of Tech., July 28, 1967).
6. M. FRIEDMAN, "The Monetary Studies of the National Bureau," *Annual Report*, 1965, Table II-1, p. 15.
7. M. FRIEDMAN AND A. J. SCHWARTZ, "Money and Business Cycles," *Rev. of Econ and Statis.*, Sup., Feb., 1963, pp. 32–64.
8. D. LAIDLER, "The Permanent-Income Concept in a Macro-Economic Model," *Oxford Econ. Papers*, Mar., 1968, pp. 11–23.
9. F. DE LEEUW AND E. GRAMLICH, "The Federal Reserve-MIT Econometric Model," *Fed. Res. Bul.*, Jan., 1968, pp. 11–40.
10. D. P. TUCKER, "Income Adjustment to Money-Supply Changes," *A.E.R.*, June, 1966, pp. 431–49.

APPENDIX TABLE

REGRESSION OF COMMERCIAL PAPER RATE ON LAGGED VALUES OF MONETARY GROWTH RATE, MONTHLY DATA 1910–65, FIRST DIFFERENCES

Regression Cofficients		t Values	Cumulative Sum of β Coefficients
Constant	2.2	.03	
β_0	−0.35	−5.5	−0.35
β_1	−0.57	−6.8	−0.92
β_2	−0.60	−5.9	−1.52
β_3	−0.50	−4.5	−2.02
β_4	−0.33	−2.8	−2.35
β_5	−0.22	−1.8	−2.57
β_6	−0.06	−0.5	−2.63
β_7	0.07	0.6	−2.56
β_8	0.22	1.8	−2.34
β_9	0.28	2.3	−2.06
β_{10}	0.28	2.3	−1.78
β_{11}	0.34	2.8	−1.44
β_{12}	0.39	3.2	−1.05
β_{13}	0.37	3.0	−0.68
β_{14}	0.34	2.7	−0.34
β_{15}	0.43	3.5	+0.09
β_{16}	0.41	3.4	+0.50
β_{17}	0.37	3.0	+0.87
β_{18}	0.40	3.3	+1.27
β_{19}	0.41	3.3	+1.68
β_{20}	0.43	3.5	+2.11
β_{21}	0.38	3.1	+2.49
β_{22}	0.32	2.6	+2.81
β_{23}	0.19	1.5	+3.00
β_{24}	0.12	1.0	+3.12
β_{25}	0.06	0.5	+3.18
β_{26}	−0.01	−0.1	+3.17
β_{27}	−0.03	−0.2	+3.14
β_{28}	−0.07	−0.6	+3.07
β_{29}	−0.14	−1.2	+2.94
β_{30}	−0.11	−0.9	+2.83
β_{31}	−0.11	−0.9	+2.72
β_{32}	−0.11	−1.0	+2.61
β_{33}	−0.09	−0.8	+2.52
β_{34}	−0.15	−1.3	+2.37
β_{35}	−0.14	−1.4	+2.23
β_{36}	−0.09	−0.9	+2.12
β_{37}	−0.05	−0.5	+2.07
β_{38}	−0.06	−1.0	+2.01

NOTE: For the form of the regression equation, see text equation in Section III. Units of the coefficients are basis points per 1 percentage point change in monthly rate of change of money stock.

$R = .367$

$\text{Adj.}R^2 = .081$

MONEY AND PERMANENT INCOME: SOME EMPIRICAL TESTS*

By James Tobin *and* Craig Swan
Yale University

According to an increasingly influential school of thought, centered in this city, variation in the money supply is the principal determinant—indeed virtually the exclusive determinant—of variation in money income. The supporting arguments have been more empirical than theoretical. The empirical evidence has included careful historical narrative [6], systematic investigation of cyclical leads and lags among relevant time series [7], and single-equation regression analyses [4]. Less attention has been given to the task of providing a theoretical rationale of the empirical findings, a monetary theory of income determination to set against the neo-Keynesian models of many macroeconomics textbooks. However, Friedman and Schwartz (FS) have presented an explicit model in [7]. Their "permanent income" theory of money demand has testable implications, and in this paper we test some of them.

The Friedman-Schwartz Model

First, a brief outline of the permanent income theory: FS hypothesize that per capita demand for real money balances is related to permanent real income as follows:

$$M/P_p = \gamma y_p^{\delta} \tag{1}$$

where

M = nominal money stock (Currency + Demand Deposits + Time Deposits), per capita

P_p = permanent price index of consumer goods

y_p = permanent real income, per capita

The money stock M is taken to be exogenous; demand must adapt to the supply. Equation (1) is, as a first approximation, always satisfied; the economy is always on its money demand curve.

The permanent value of a variable—price or income—is a weighted average of its current and past actual values, with account taken of trend. For a variable, $X(t)$, permanent $X_p(t)$ is defined as follows:

$$X_p(t) = (1+\alpha_X)^t \prod_{i=0}^{\infty} [X(t-i)/(1+\alpha_X)^{t-i}]^{w_i} \tag{2}$$

where

α_X is the trend rate of growth of X

w_i is the exponential weight of actual X_{t-i};

$$\sum_{i=0}^{\infty} w_i = 1.$$

With several substitutions and simplifications, (2) can be expressed in a more manageable form. First, taking the logarithms of (2) yields

$$\log X_p(t) = t \log [1+\alpha_X] + \sum_{i=0}^{\infty} w_i\{\log X(t-i) - (t-i)\log(1+\alpha_X)\}$$

which can be rewritten as:

$$\log X_p(t) - t\log(1+\alpha_X) = \sum_{i=0}^{\infty} w_i\{\log X(t-i) - (t-i)\log(1+\alpha_X)\} \tag{3}$$

Second, FS assume that the w_i can be characterized as a simple geometrically declining series of weights which sum to unity; i.e., $w_i = w_o(1-w_o)^i$. It is convenient to note that

---

* The research described in this paper was carried out under grants from the National Science Foundation and from the Ford Foundation.

$$\sum_{i=1}^{\infty} w_o(1-w_o)^i = (1-w_o)\sum_{j=0}^{\infty} w_o(1-w_o)^j \tag{4}$$

This property of the lag structure makes possible the following substitution in (3):

$$\begin{aligned}\log X_p(t) - t\log(1+\alpha_X) &= w_o[\log X(t) - t\log(1+\alpha_X)] \\ &\quad + (1-w_o)[\log X_p(t-1) \\ &\quad - (t-1)\log(1+\alpha_X)]\end{aligned}$$

or

$$\begin{aligned}\log X_p(t) &= w_o\log X(t) \\ &\quad + (1-w_o)\log X_p(t-1) \\ &\quad + (1-w_o)\log(1+\alpha_X)\end{aligned} \tag{5}$$

Taking logarithms of equation (1), the equality of supply and demand for real money balances, gives

$$\begin{aligned}\log M(t) &= \log\gamma + \delta\log y_p(t) \\ &\quad + \log P_p(t).\end{aligned} \tag{6}$$

FS assume that the w_i associated with y_p and P_p are identical. Using (5) to calculate y_p and P_p gives:

$$\begin{aligned}\log M(t) &= \log\gamma + \delta w_o\log y(t) \\ &\quad + w_o\log P(t) \\ &\quad + (1-w_o)[\log(1+\alpha_P) \\ &\quad + \delta\log(1+\alpha_y)] \\ &\quad + (1-w_o)[\delta\log y_p(t-1) \\ &\quad + \log P_p(t-1)]\end{aligned} \tag{7}$$

Since FS assume that the economy is always on its demand curve for money, they can use (6) for period $t-1$ to eliminate the unobserved permanent variables in (7) and get

$$\begin{aligned}\log M(t) &= w_o\log\gamma + (1-w_o) \\ &\quad [\log(1+\alpha_P) + \delta\log(1+\alpha_y)] \\ &\quad + \delta w_o\log y(t) + w_o\log P(t) \\ &\quad + (1-w_o)\log M(t-1)\end{aligned} \tag{8}$$

Equation (8) can then be solved for $\log y(t)$:

$$\begin{aligned}\log y(t) &= -\frac{1}{\delta}\log\gamma - \frac{(1-w_o)}{\delta w_o} \\ &\quad [\log(1+\alpha_P) + \delta\log(1+\alpha_y)] \\ &\quad + \frac{1}{\delta w_o}[\log M(t) - \log P(t)] \\ &\quad - \frac{1-w_o}{\delta w_o}[\log M(t-1) - \log P(t)].\end{aligned} \tag{9}$$

Equation (9) is an expression for real per capita income. It could be converted into an expression for money per capita income by adding $\log P(t)$ to both sides.[1] This would leave $\log P(t)$ on the right with a coefficient of $1-1/\delta$. For purposes of estimation, equation (9) has the advantage of being identified. It can be used to derive estimates of δ, the elasticity of money demand with respect to permanent income, and w_o, the weight of current information in estimating permanent values. An equation for money income would be over-

---

[1] However, there are problems in using equation (9). First, as indicated below the price level is not really exogenous; there is another structural relationship concerning the division of increases in aggregate demand between real income and prices. Single equation treatment of equation (9) ignores this other relationship. Second, it is by no means obvious that the stochastic elements in the model produce a well-behaved additive error term in equation (9). Suppose that (1) were $M/P_p = \gamma y_p^{\delta}\cdot\exp^{\epsilon}$ where ϵ is a normally distributed error. The error in (9), call it $\eta(t)$, will then be $\epsilon(t) - (1-w_0)\epsilon(t-1)$. For $\eta(t)$ to be serially independent, $\epsilon(t)$ would have to be positively serially correlated in a specific manner. If $\epsilon(t)$ is serially independent, then $\eta(t)$ will show negative serial correlation (not the high positive serial correlation shown in the residuals from the level regression).

TABLE 1

ANNUAL AND QUARTERLY REGRESSIONS OF INCOME ON MONEY 1951–66

Equation	Intercept	Coefficient of:*		R^2	DW	n	w_o	δ
		$\log \frac{M(t)}{P(t)}$	$\log \frac{M(t-1)}{P(t)}$					
Annual								
Levels..........	.74	1.43 (.530)	−.61 (.709)	.76	.74	16	.57	1.23
First differences..		.27 (.320)	.62 (.364)	.32	2.44	15	3.30	1.12
Quarterly								
Levels..........	.73	1.31 (.707)	−.35 (.743)	.74	.11	64	.73	1.04
First differences..		.12 (.239)	.61 (.238)	.25	1.18	63	6.08	1.37

* Standard errors of the coefficients are reported in parentheses.
R^2 = Coefficient of determination
DW = Durbin-Watson statistic
n = number of observations
For data sources see Data Appendix.

identified. Although equation (9) concerns real income, there is no implication that the effects of an increase in the money supply will affect real income to the exclusion of prices. Actually FS, like other economists, expect short-run changes in money income to be divided between output and prices.

Estimates from Annual and Quarterly Regressions 1951–66

Equation (9) was fitted to annual and quarterly data from 1951 through 1966. Results of level and first difference forms of the regression are reported in Table 1.

On the basis of his study of the consumption function and other work, Friedman estimates the weight of current year income in permanent income at $\frac{1}{3}$. (This would imply a value of w_o of .096 in the quarterly regressions.) Friedman also estimates δ, the elasticity of money demand with respect to permanent income, to be 1.8 (see [7]).

Our estimates of w_o are higher than the estimates Friedman has reported and our estimates of δ are considerably lower than Friedman's own in [1]. Friedman's estimates referred to a longer time period. Virtue does not necessarily lie with long time periods; structural changes have occurred. Commercial banks have in recent decades faced much stronger competition for savings from other financial intermediaries than they did in the late nineteenth and early twentieth century. (See [8, p. 105].) It is not surprising, therefore, that Friedman's estimates of δ, the long-run income elasticity of money demand, are higher than ours. Furthermore, the major financial reforms of the 1930's might well have changed these parameters.

The Durbin-Watson statistics suggest that the level regressions show high positive serial correlation of the residuals. Use of first differences meets this problem, but the explanatory power of the model then becomes very low, and in terms of the theory the estimates of w_o become absurd. (Indeed, they suggest an opposite model, in which permanent money holdings are related to current income.) A tempting interpretation is that the correlation exhibited in the level regressions reflects common trends in income and money rather than a causal relationship between the variables, and that there are important, serially persistent nonmonetary determinants of income.

A low value of w_o plays an important part in FS's explanation of the observation that short-run fluctuations in income are larger in amplitude than the monetary fluctuations that cause them. The income velocity of money moves pro-cyclically, and the permanent income model is supposed to explain this fact, among others. When money supply increases exogenously, faster than the permanent price level, permanent income must increase sufficiently to absorb the new money. If δ is 1.8, as FS estimate, then permanent income must rise .55 percent to create demand for an addition of 1 percent to the real stock of money. But the only component of permanent income that can rise is current income; the past incomes that enter the weighted average are irrevocably fixed. With $w_o=\frac{1}{3}$ for annual incomes, current year's income must rise 3 percent to raise permanent income 1 percent, or 1.65 percent to raise permanent income the necessary .55 percent. The calculation illustrates how the model reconciles FS's finding that velocity declines in the long run ($\delta>1$) with their finding that short-run changes in money stock cause more than proportionate changes in income.

FS recognize in [7] that the model proves too much if it is applied literally to quarterly data. As already noted, the $\frac{1}{3}$ estimate for w_o for annual data implies a weight of .096 for the income of the current quarter. If the entire income adjustment to a change in money stock must occur within a quarter, then it will be more than three times as large as indicated in the previous paragraph. It will take a 5.7 percent rise in current income for the quarter to raise permanent income .55 percent. As Table 1 indicates, our quarterly regressions indicate much larger values of w_o than the FS model, literally applied, would imply. As FS themselves suggest, perhaps we should relax the assumption that money demand adjusts so rapidly as to keep the community on its demand curve every quarter.

In this spirit we introduce the following modification: Assume people do not adjust their current money balances to income and prices but rather adjust a weighted average of the current and preceding quarters' money balances. Consequently we define $M^*(t)$ as $(1-\beta)M(t)+\beta M(t-1)$ and recompute the quarterly regressions with $M^*(t)$ substituted for $M(t)$. Results for $\beta=.25$, .5, .75 and 1.0 are reported in Table 2. The quarterly results in Table 1 are equivalent to $\beta=0$. While estimates of w_o decline as β rises, they are still large. They imply a much larger response of the demand for money to changes in current income than the FS model.

Interest Rate Effects?

An alternative explanation of observed pro-cyclical movements of velocity is sensitivity of money demand to interest rates. Given such sensitivity, short-run fluctuations in income can have nonmonetary as well as monetary causes. If the monetary authorities "lean against the wind," then money supply, interest rates, and velocity will all increase in booms and decline in recessions. This would be a Keynesian interpretation of the same observations that the FS model is designed to explain. It would of course have very different policy implications, leaving room for fiscal policy and exogenous changes in

TABLE 2

QUARTERLY REGRESSIONS OF INCOME ON CURRENT AND LAGGED MONEY STOCKS
1951.1–1966.4

Equation	Intercept	Coefficient of: $\log\left(\frac{M^*(t)}{P(t)}\right)$	Coefficient of: $\log\left(\frac{M^*(t-1)}{P(t)}\right)$	R^2	DW	w_0	δ
Levels							
$\beta=.25$	.73	1.69 (.778)	− .74 (.821)	.74	.11	.56	1.05
$\beta=.5$	.73	2.03 (.817)	−1.09 (.865)	.74	.11	.46	1.06
$\beta=.75$	.73	2.20 (.806)	−1.26 (.856)	.74	.12	.43	1.06
$\beta=1.0$	.73	2.14	−1.20	.74	.15	.44	1.06
First differences							
$\beta=.25$		.01 (.313)	.76 (.311)	.28	1.21	77.0	1.30
$\beta=.5$		−.01 (.365)	.82 (.364)	.29	1.21	−81.0	1.23
$\beta=.75$		.20 (.311)	.62 (.310)	.29	1.20	4.10	1.21
$\beta=1.0$		.34 (.233)	.47 (.233)	.29	1.21	2.38	1.23

private spending, as well as monetary events, to affect income.

While Friedman has doubted the empirical significance of interest rates, other than expected changes in the value of money, on the demand for money [1], other researchers have found evidence of such influence. (See, for example, [9] [10].) Certainly there is ample theoretical reason to suppose that the holding of money is influenced by its own real yield—which for the FS definition of money depends both on the rate paid by commercial banks on time deposits and on the expected rate of price change—and by the real yields on substitute assets such as Treasury bills.

These effects can be built into the FS equation (1) as follows:

$$\frac{M(t)}{P_p(t)} = \gamma y_p(t)^{\delta}\left(\frac{P_p(t)}{P_p(t-1)}\right)^{v} R_{TD}(t)^{a} R_T(t)^{c} \tag{10}$$

where

R_{TD} = rate paid by commercial banks on time deposits,

R_T = market yield on new issue Treasury bills.

$$\left(\frac{P_p(t)}{P_p(t-1)}\right)$$

is a measure of the change in the permanent price level; it is also, as can be shown by application of equation (5) letting X

TABLE 3

ANNUAL AND QUARTERLY REGRESSIONS OF INCOME ON MONEY AND RATES OF RETURN

	Intercept	Coefficient of:			
		$\log \frac{M(t)}{P(t)} - \log \frac{M(t-1)}{P(t)}$	$\log P(t) - \log P(t-1)$	$\log R_{TD}(t) - \log R_{TD}(t-1)$	$\log R_T(t)$ $-\log R_T(t-1)$
Annual	.70	.66 (.050)	−.18 (.230)	.05 (.012)	.05 (.013)
Quarterly	.70	.66 (.025)	−.29 (.339)	.05 (.005)	.05 (.005)

	R^2	DW	w_0	δ	ν	η	ϵ
Annual	.98	1.62	.99	1.52	.27	−.08	−.08
Quarterly	.98	.68	.97	1.52	.44	−.08	−.08

ν Elasticity of demand for real money balances with respect to permanent price increase.
η Elasticity of demand for real money balances with respect to the rate paid on time deposits.
ϵ Elasticity of demand for real money balances with respect to the yield on new issue Treasury bills.

equal $P(t)/P(t-1)$, the permanent value of price change. Substitutions similar to those previously made in equation (1) yield the following formulation:

$$
\begin{aligned}
\log y(t) = {} & -\frac{1}{\delta}\log\gamma - \frac{1-w_o}{\delta w_o} \\
& \cdot[\log \alpha_P + \delta \log \alpha_\nu] \\
& - \frac{\nu}{\delta}[\log P(t) - \log P(t-1)] \\
& + \frac{1}{\delta w_o}\{[\log M(t) - \log P(t)] \\
& - (1-w_o)[\log M(t-1) \\
& - \log P(t)]\} \\
& - \frac{\eta}{\delta w_o}\{\log R_{TD}(t) \\
& - (1-w_o)\log R_{TD}(t-1)\} \\
& - \frac{\epsilon}{\delta w_o}\{\log R_t(t) \\
& - (1-w_o)\log R_t(t-1)\}
\end{aligned}
\tag{11}
$$

It is easily seen that equation (11) is overidentified; aside from constants, there are seven coefficients to determine five parameters. But by fixing w_o it is possible to combine terms for $M(t)$ and $M(t-1)$, $R_{TD}(t)$ and $R_{TD}(t-1)$, and $R_T(t)$ and $R_T(t-1)$ and to regard (11) as an equation involving four coefficients to determine four parameters other than w_o. Equation (11) was then estimated by varying w_o in steps between 0 and 1.0 and choosing that value of w_o which maximized the R^2.[2] Results are presented in Table 3.

The estimates of δ in this more general version, equation (10), are larger than before. Perhaps the exclusion of relevant interest rate terms biased downward the earlier estimates of δ. High income may have served as a proxy for high market interest rates, which have a predominantly negative effect on money holdings. More surprising is the positive coefficient on the R_{TD} term. This result may be in part explained by the strong correlation between the time deposit rate and the Treasury bill rate and by shifts within M

---

[2] In fact w_0 was first varied from .1 to .9 in increments of .1. w_0 was then varied by .01 in the region of the previous maximum.

between demand and time deposits. Most surprising, however, is the apparent positive elasticity of money holdings with respect to price changes. Perhaps people do not in fact extrapolate current and past price trends. Finally, these results do not confirm the hypothesis that behavior will be better explained by relating it to permanent values of income, price, and price change. The estimates of w_e are so close to one as to eliminate almost all difference of permanent from current values.

Prediction Tests

Another test of the basic structure of equation (1) is to see how good a predictor it is. Friedman has expressed the view that "the only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience" [2, pp. 8–9]. Correspondingly, a variant of (9) was used to predict annual and quarterly changes in money income from 1959 through 1968. Expressing $\log y(t)$ as $\log Y(t) - \log P(t)$ and taking first difference of (9) yields:

$$\begin{aligned} \Delta \log Y(t) &= \frac{1}{\delta w_e} \Delta \log M(t) \\ &\quad - \frac{1 - w_e}{\delta w_e} \Delta \log M(t-1) \\ &\quad - \frac{w_e(1-\delta)}{\delta w_e} \Delta \log P(t) \end{aligned} \tag{12}$$

Let each first difference be expressed as a deviation from its average value; if X is subject to a geometric trend $X(0)(1+\alpha_X)^t$, this contributes a constant, $\log(1+\alpha_X)$, to growth of $\log X$ each period. Thus, let

$$\begin{aligned} \tilde{Y}(t) &= \Delta \log Y(t) - \log(1 + \alpha_Y) \\ \tilde{P}(t) &= \Delta \log P(t) - \log(1 + \alpha_P) \\ \tilde{M}(t) &= \Delta \log M(t) \\ &\quad - \log[1 + \delta(\alpha_Y - \alpha_P) + \alpha_P] \end{aligned}$$

For the present tests, α_Y and α_p were measured as the actual average compound rates of change of money income and prices from 1950 to 1960. Substituting these definitions into (12) yields

TABLE 4

PREDICTIONS OF ANNUAL PERCENTAGE CHANGES IN MONEY INCOME 1959–67

Year	Actual	FS13	FS14	Naïve*	Trend†	Adjusted Trend‡
59	8.3	3.2	3.6	1.1	5.6	5.5
60	4.1	−2.9	−3.1	8.3	5.6	5.8
61	3.2	10.8	11.8	4.1	5.6	5.6
62	7.5	6.8	7.6	3.2	5.6	5.4
63	4.4	7.0	7.6	7.5	5.6	5.6
64	7.0	5.2	5.7	5.4	5.6	5.6
65	8.0	6.1	6.6	7.0	5.6	5.7
66	8.8	8.1	8.2	8.0	5.6	5.8
67	6.6	5.4	4.7	8.8	5.6	6.0
Average absolute change...	8.7					
Average absolute error.....		3.1	3.2	2.7	1.9	1.8

* Actual change last period taken as forecast.

† Average annual rate of change, 1950–60.

‡ Predicted change for year t is found by solving for r
$$Y(t-1) = (1+r)^{(t-1-1950)} Y(1950).$$

TABLE 5

PREDICTIONS OF QUARTERLY PERCENTAGE CHANGE IN MONEY INCOME
1959.1–1968.2

Quarter	Actual	FS13	FS14	Naïve*	Trend†	Adjusted Trend‡
59.1	2.1	1.1	1.1	3.1	1.6	1.6
2	2.7	−.7	−.7	2.1	1.6	1.6
3	−.8	−1.1	−1.2	2.7	1.6	1.6
4	1.2	−5.4	−5.8	−.8	1.6	1.5
60.1	2.6	−.6	−.7	1.2	1.6	1.5
2	.4	3.1	3.4	2.6	1.6	1.5
3	−.1	8.2	9.1	.4	1.6	1.5
4	−.2	1.4	1.5	−.1	1.6	1.5
61.1	−.1	2.5	2.7	−.2	1.6	1.4
2	2.3	1.9	2.2	−.1	1.6	1.4
3	1.8	.8	1.0	2.3	1.6	1.4
4	2.7	1.9	2.0	1.8	1.6	1.4
62.1	1.5	3.6	4.0	2.7	1.6	1.5
2	1.8	.1	.3	1.5	1.6	1.5
3	1.3	−5.7	−6.1	1.8	1.6	1.5
4	1.4	12.0	13.0	1.3	1.6	1.5
63.1	.9	.1	.1	1.4	1.6	1.5
2	1.1	1.0	1.1	.9	1.6	1.5
3	1.9	.4	.5	1.1	1.6	1.5
4	1.9	2.4	2.5	1.9	1.6	1.5
64.1	1.8	−.5	−.6	1.9	1.6	1.5
2	1.8	−1.5	−1.6	1.8	1.6	1.5
3	1.6	8.5	9.4	1.8	1.6	1.5
4	.9	−.6	−.7	1.6	1.6	1.5
65.1	2.7	2.9	3.1	.9	1.6	1.5
2	1.8	−.6	−.8	2.7	1.6	1.5
3	2.0	−.7	−.7	1.8	1.6	1.5
4	2.8	13.9	15.3	2.0	1.6	1.5
66.1	2.5	−7.2	−8.0	2.8	1.6	1.5
2	1.5	2.6	2.6	2.5	1.6	1.5
3	1.8	−4.8	−5.4	1.5	1.6	1.5
4	1.9	−2.9	−3.3	1.8	1.6	1.5
67.1	.4	12.4	13.5	1.9	1.6	1.5
2	.9	5.7	6.1	.4	1.6	1.5
3	1.9	3.8	3.9	.9	1.6	1.5
4	2.0	−3.1	−3.7	1.9	1.6	1.5
68.1	2.6	−3.0	−3.5	2.0	1.6	1.5
2	2.4	2.0	1.8	2.6	1.6	1.5
Average absolute change...	1.6					
Average absolute error.....		3.7	3.9	.8	.7	.7

* Actual change last period taken as forecast.

† Average annual rate of change, 1950–60.

‡ Predicted change for year t is found by solving for r

$$Y(t-1) = (1+r)^{(t-1-1950)}Y(1950).$$

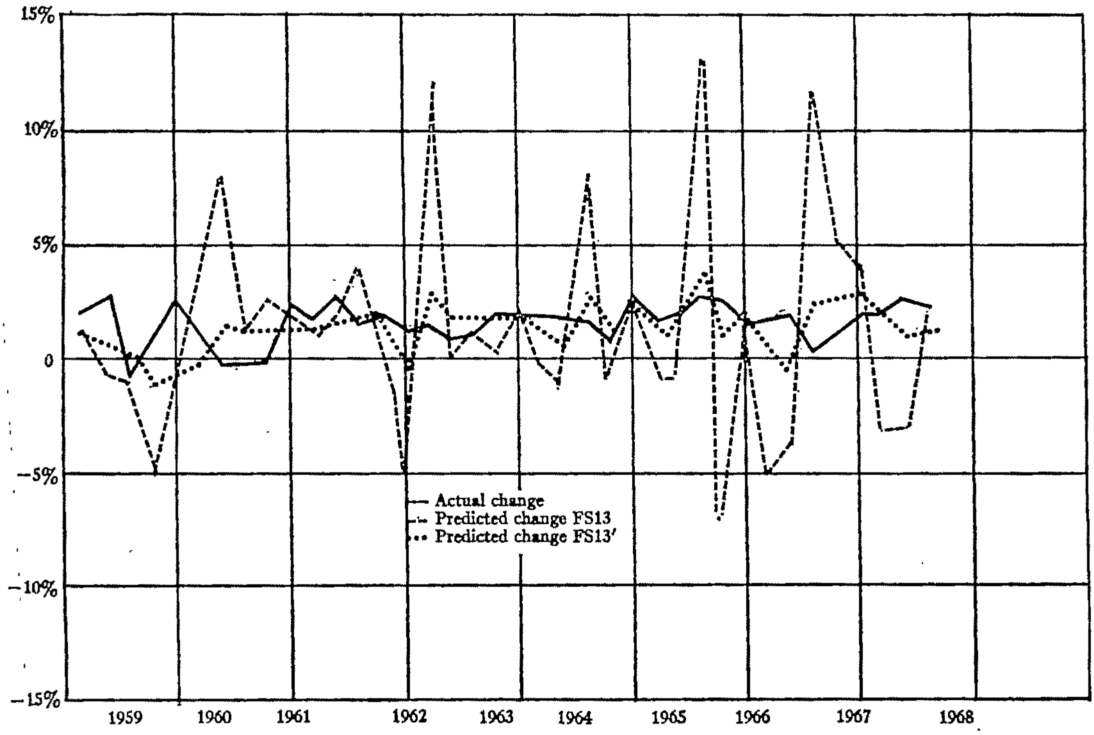


FIGURE 1

QUARTERLY PERCENTAGE CHANGES IN MONEY INCOME 1959.1–1968.2

$$\tilde{Y}(t) = \frac{1}{\delta w_o} \tilde{M}(t) - \frac{1 - w_o}{\delta w_o} \tilde{M}(t-1) - \frac{w_o(1-\delta)}{\delta w_o} \tilde{P}(t) \tag{13}$$

Use of (13) yields predictions of deviations of money income per capita from its trend. Predictions of actual percentage changes in aggregate money income can then be obtained by adding population change and trend change in income. FS estimates of w_o and δ, .33 and 1.81, respectively, were used. For quarterly predictions w_o was set equal to .096.

In [7] FS indicate that one might expect prices and money income to move together systematically. They consider the elasticity of the measured price level with respect to measured income and assign it a value of .2. Substituting .2Δ log $Y(t)$ for Δ log $P(t)$ in (12) yields another predictor of changes in money income,

$$\tilde{Y}(t) = \frac{1}{(.2 + .8\delta)w_o} \tilde{M}(t) - \frac{1 - w_o}{(.2 + .8\delta)w_o} \tilde{M}(t-1) \tag{14}$$

Results of these predictions are presented in Tables 4 and 5. FS13 indicates predictions based on equation (13) and FS14 indicates predictions based on equation (14). Three simpleminded modes of prediction are presented for comparison, in the last three columns of Tables 4 and 5. Figure 1 illustrates actual quarterly percentage changes in money income and the predictions of those changes based on equation (13). The dashed line, FS13 is a graph of the figures reported in Table 5.

The dotted line FS13′ uses estimates of w_o and δ—.73 and 1.04—from our original quarterly regression (see Table 1).

A theory that leads to worse error than the naïve hypothesis that last year repeats itself is of questionable reliability. The quarterly predictions are even worse than the annual; see Table 5 and Figure 1. This may be partly due to the problems of lag structure discussed above. As our own experiment indicated, quarterly results can be somewhat improved by relaxing the requirement that current income adjust enough to create demand for all the new money, requiring instead that only a fraction of the new money supply be immediately matched by permanent new demand. This can be done, but only at some expense to the power of the FS hypothesis to produce a large money multiplier and to explain how short-run variations in money creation induce reinforcing changes in velocity.

The moral of the exercises of estimation and prediction that we have presented here is simple. Contrary, perhaps, to much popular belief, the evidence does not support the view that there is a simple, direct relationship of income to money. Policymakers and forecasters would not have much luck in trying to infer movements of money income from changes in money stock. The permanent income hypothesis is an interesting theoretical rationale for certain qualitative features of observed fluctuation of income and money. But it does not fit postwar data very well, and our results certainly provide no reason to prefer the FS model to a Keynesian interest rate interpretation of short-run fluctuations in the demand for money.

DATA APPENDIX

M Demand deposits+time deposits+currency per capita (thousands)
From: *Fed. Res. Bul.*, June, 1964; Sept., 1966,
and *Survey of Current Business*, Jan., 1967.

Y Net National Product per capita (thousands of current dollars)
From: *National Income and Product Accounts of the U.S. 1929–1965*, Table 1.9, pp. 12–13
and *Survey of Current Business*, Jan., 1967.

P GNP deflator (1958=100)
From: *National Income and Product Accounts of the U.S. 1929–1965*, Table 8.1, pp. 158–59
and *Survey of Current Business*, Jan., 1967.

POP Total Population (millions)
From: *Current Population Reports, Population Estimates*, Census Series P-25
No. 331 Mar. 22, 1966
No. 357 Jan. 18, 1967
Population for year defined as population on July 1.
Population for quarter defined as average of population for first day of second and third month of quarter.

y Net National Product per capita (constant 1958 dollars)
Y/P

R_{TD} Rate paid by commercial banks on time deposits.
From: U.S. Savings and Loan League *Fact Book*, 1968.
Published figures assumed to represent rate paid on June 30.
Quarterly figures from linear interpolation to mid-quarter.

R_T Market yield on new issue 3-month Treasury bills
From: *Business Statistics*, 1967 (pp. 90 and 237)

Data for predictions through 1968.2 came from the *Survey of Current Business*, July, 1968, or from more recent sources as listed above. In order to eliminate the problem of data revisions, values for 66.4 were altered in proportion to the more recent data.

REFERENCES

1. M. FRIEDMAN, "The Demand for Money: Some Theoretical and Empirical Results," *J.P.E.*, Aug., 1959, pp. 327–51.
2. ———, *Essays in Positive Economics* (Univ. of Chicago Press, 1953).
3. ———, *Studies in the Quantity Theory of Money* (Univ. of Chicago Press, 1956).
4. ———, *A Theory of the Consumption Function* (N.B.E.R., Princeton Univ. Press, 1957).
5. M. FRIEDMAN AND D. MEISELMAN, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in CMC *Stabilization Policies* (Prentice-Hall, Inc., 1963).
6. M. FRIEDMAN AND A. SCHWARTZ, *A Monetary History of the United States* (Princeton Univ. Press, 1963).
7. M. FRIEDMAN, "Money and Business Cycles," *Rev. of Econ. and Statis.*, Feb., 1963, Sup., pp. 32–64.
8. M. J. HAMBURGER, "Household Demand for Financial Assets," *Econometrica*, Jan., 1968, pp. 97–118.
9. H. R. HELLER, "The Demand for Money: The Evidence from Short Run Data," *Q. J. E.*, May, 1965, pp. 291–303.
10. H. A. LATANÉ, "Cash Balances and the Interest Rate: A Pragmatic Approach," *Rev. of Econ. and Statis.*, Nov., 1954, pp. 456–60.

ECONOMETRIC ANALYSIS OF STABILIZATION POLICIES*

By ALBERT ANDO, *University of Pennsylvania*
and
FRANCO MODIGLIANI, *Massachusetts Institute of Technology*

I. *Purpose and Approach*

The purpose of this paper is to provide an interim report on the performance of the FRB-MIT quarterly econometric model of the postwar American economy and of its use in the analysis of stabilization policies. The model is being developed jointly by a team of academic economists under the direction of the authors and by economists of the Board of Governors of the Federal Reserve System.

Since the publications of some preliminary reports a year ago [6] [18] extensive revisions have been undertaken of certain sectors—notably, consumption and housing—and some sectors which were treated as exogenous have been endogenized—notably, the labor market and the price wage sectors. Even at this stage, however, we regard our results as preliminary and tentative. Accordingly, our emphasis here will be more on the methodological problems involved in the analysis rather than on specific quantitative conclusions.

Due to limitations of space we do not present a complete listing of the equations,[1] nor do we attempt a detailed explanation of the structure, which is better provided in a number of papers already completed or forthcoming. We merely present in Table 1 a bird's eye view of the structure of the model by listing its major sectors and subsectors and indicating, for each, the number of "behavioral" equations, i.e., those involving empirically estimated parameters, and the number of estimated parameters. (In addition, the model includes a generous number of definitional and accounting identities, which brings the number of equations and endogenous variables to close to 200.) In the course of the exposition we will also provide brief sketches of some of the most significant innovations that distinguish our model from other existing ones. One such feature directly related to the focus of our research is that roughly a half of behavioral equations are concerned with fiscal and financial variables.

The approach to the econometric analysis of stabilization policies on which we rely in this paper differs somewhat from that pursued by the very small number of predecessors [1] [8] [7] and needs to be outlined briefly.

Broadly speaking, the final or at least intermediate goal of our research is that of analyzing and measuring the effects of changes in policy variables on certain key economic variables. These effects may be visualized as occurring through a variety of channels or mechanisms, each of which may be more or less well understood and reliably measured in a given state of the arts, and hence more or less accurately predictable. In this paper we propose to identify and analyze a number of such basic mechanisms and to assess how reliably our model succeeds in capturing their workings, both separately and jointly.

We begin in Section II with an analysis of the "pure consumption multiplier"

---

* The authors gratefully acknowledge their indebtedness to Professor Robert Rasche, of the University of Pennsylvania, who took major responsibility for the computational and organizational aspects of getting the system to the operational stage, and to Laurence H. Meyer, of M.I.T., and Abram Hoffman, of the Univ. of Pennsylvania, for carrying out the simulations reported in the paper.

[1] Such a listing will be available from the authors on request.

mechanism, which results from the interaction of consumption, output, and private income, but excludes all feedbacks through other components of aggregate demand, or through financial markets and price effects. In the context of this "limited feedback system" we examine the multiplier effects of a change in government expenditure. In the next three sections we successively allow for other feedback channels, occurring through investment (multiplier-accelerator interaction), through financial markets and through wage-price effects, at each step assessing their reliability, as well as their effects on the size and time path of the government expenditure multiplier.

In carrying out the program outlined in the paragraph above, we make extensive use of dynamic simulations of sectors, of subsectors, and of the whole model. In these dynamic simulations, we take as given the actual time path of a number of variables, which can be usefully grouped into four classes: (1) Exogenous variables; i.e., variables that we do not explain in the context of our model. These include mainly variables such as population and its structure that are deemed to be unaffected by the economic process, at least in the short run (technically, to be uncorrelated with the residual errors of the behavior equations). But they include also certain variables that are only provisionally taken as given, pending completion of further work. The important ones are exports, relative prices, most capital gains, variables related to the farm sector and a few relatively "minor" variables which will be listed below. (2) Policy variables which are under the control of the policy-makers. In our model these include both fiscal variables—federal government tax and transfer rates, expenditures on goods and services and grants-in-aid to state and local governments—and monetary variables—unborrowed bank reserves and reserve requirements, discount rate and ceilings on some interest rates. Although in our simulations we take these policy variables as given, in a broader context they may be responding to the unfolding of the economy through the discretion of policy-makers.[2] (3) Experimentally exogenous variables; i.e., variables that are endogenous to the full model, but are taken as given in a particular simulation for the purpose of analyzing the performance and implications of specific sectors or "limited feedback" systems. (4) The past history of endogenous variables up to the date at which the dynamic simulations begin. After that date all the endogenous variables, current and lagged, are generated by simultaneous solution of the equations of the sector or system.

The basic measure of reliability of any sector in this paper will be the root mean square deviation (RMS) of the actual path of the endogenous variables from the path generated in a fairly long (some forty to fifty quarters) dynamic simulation of the sector. The simulation errors so measured, in contrast to residual errors of estimation for individual equations, are of critical importance in evaluating the model. This is because in the process of simulation solutions at each point of time and over time (through lagged dependent variables) there will be an opportunity for the single equations' errors to interact with each other and to be magnified as they pass through other equations of the system. In principle, even modest single equation errors could build up into progressively larger errors, causing the

---

[2] This possibility, of course, poses some serious estimation problems. Similar difficulty is encountered when two or more potential policy variables are related with each other through equations in the model. In the context of our model, for example, the central bank could conceivably control unborrowed reserves or the short-term interest rate, or sometimes one and sometimes the other, but never both simultaneously.

TABLE 1

MEASURES OF PERFORMANCE OF THE MODEL

(The top figure in each box is RMS, root mean squares of the difference between actual and simulated values; the bottom figure, italicized, is RMS divided by the average actual values)

Sectors			Final Demand									
Subsectors			Consumption			S&L Govt.	Imports	Inventory	Investment in Plant & Equipment			Housing
Number of Equations			3			3	1	1	10			7
Number of Parameters			34			44	5	15	55			67
Row	System and Period	Major Variables / Experimentally Exogenous Variables	Consumption '58 $, bill. (1)	Consumers' Durables '58 $, bill. (2)	Consumers' Expenditure '58 $, bill. (3)	State & Local Government Exp. $, bill. (4)	Imports '58 $, bill. (5)	Inventory '58 $, bill (6)	New Orders for Equipment '58 $, bill. (7)	Equipment Expenditure '58 $, bill. (8)	Plant Expenditure '58 $, bill. (9)	Residential Construction $, bill. (10)
(a)	Subsector* 1955.1, 67.2	(See text)	1.4 *0.42*	1.5 *2.7*	2.1 *0.61*	1.4 *2.6*	1.5† *4.9*	2.9 *61.* *56.§*	1.5 *3.8*	0.87 *2.4*	0.61 *3.2*	0.70‡ *2.9*
(b)	Consumption Multiplier	All GNP components ex. consumption; Prices, Wages; Financial Sector	1.5 *0.45*	1.7 *3.3*	2.4 *0.69*							
(c)	Broad Multiplier 1955.1–67.2	All Investment Prices, Wages, Financial Sector	1.5 *0.45*	1.7 *3.3*	2.3 *0.68*	1.2 *2.4*	1.7 *6.6*					
(d)	All "Real" Sectors 1956.1–67.2	Prices-Wages Financial Sector	1.8 *0.50*	1.7 *4.9*	2.4 *0.70*	1.4 *2.6*	1.6 *5.0*	3.7	2.1 *5.0*	1.6 *4.0*	0.63 *3.2*	0.76 *3.2*
(e)	Real & Financial Sectors 1956.1–67.2	Price-Wages	1.9 *0.55*	1.7 *5.0*	2.5 *0.75*	1.6 *3.0*	1.7 *6.9*	3.8	2.4 *7.1*	1.8 *5.3*	1.1 *5.9*	1.2 *4.9*
(f)	All Except Financial 1958.1–67.2	Financial Sector	2.4 *0.68*	1.8 *5.2*	2.9 *0.83*	1.4 *2.4*	2.0 *7.2*	3.9	2.4 *6.7*	1.6 *4.5*	0.70 *3.7*	0.72 *2.9*
(g)	Full System 1958.1–67.2	Capital Gains	2.4 *0.70*	2.0 *5.6*	3.1 *0.86*	1.6 *2.7*	2.2 *7.9*	4.1	3.0 *8.6*	2.0 *5.8*	0.95 *5.0*	0.97 *3.9*

computed values to depart more and more from the actual path. Fortunately, the RMS errors reported in Table 1 suggest that we have succeeded in building into our model homeostatic characteristics, which we like to think reflect those of actual economic systems, which help to keep the solution of our individual sectors as well as of the model as a whole reasonably close to the observed path, even over periods as long as fifteen years.[8]

The RMS measures of Table 1 serve in turn to calibrate the confidence that can be attached to the results of "policy simulations" designed to estimate the typical time path of response of endogenous vari-

---

[8] At the same time, it should be remembered that our measures of error undoubtedly tend to overstate the closeness of fit, because the period used in the dynamic simulation largely overlaps the period for which the single equations were estimated. This period, incidentally, is not necessarily the same for all equations. With the few exceptions noted in Table 1, it begins somewhere between 1952 and 1954, and ends in the last quarter of 1966 for those sectors which have been recently added or revised (cf. footnote 1), and in the last quarter of 1965 for the remaining equations.

TABLE 1—(*Continued*)

MEASURES OF PERFORMANCE OF THE MODEL

(The top figure in each box is RMS, root mean squares, of the difference between actual and simulated values; the bottom figure, italicized, is RMS divided by the average actual values)

Relation of GNP to Disposable Income				Labor Market			Financial Markets					Price-Wage		The Whole System		
15				3			21					2		66		
115				36			164					10		399		
Personal Income $, bill. (11)	Disposable Income $, bill. (12)	Disposable Income '58 $, bill. (13)	Deficit, Federal Government $, bill. (14)	Man-hours (bill.) (15)	Employment (mill.) (16)	Unemployment (mill.) (17)	Treasury Bill Rate (%) (18)	Corporate Bond Yield (%) (19)	Mortgage Rate (%) (20)	Dividend Yield (%) (21)	Money Supply $, bill. (22)	Price 1958=100 (23)	Wage Rate (¢) (24)	GNP $, bill. (25)	GNP '58 $, bill. (26)	Row
1.9 *0.4*	1.6 *0.43*	1.5 *0.42*	1.0	0.85 *0.70*	0.42 *0.60*	0.31 *8.6*	0.21 *7.0*	0.13 *3.2*	0.13 *2.2*	0.24 *6.9*	1.7 *1.4*	0.25 *0.25*	3.7 *1.5*			(a)
	1.8 *0.49*	1.9 *0.52*			0.45 *0.70*	0.31 *8.6*									2.4 *0.44*	(b)
		1.9 *0.51*	1.5		0.45 *0.70*	0.29 *8.5*									2.3 *0.43*	(c)
3.1 *0.69*		2.3 *0.63*	2.1	1.1 *1.0*	0.52 *0.80*	0.36 *10.0*									5.5 *1.0*	(d)
3.6 *0.84*		2.6 *0.74*	2.2	1.2 *1.07*	0.55 *0.80*	0.38 *10.1*		0.15 *3.6*	0.13 *2.2*	0.30 *8.7*					6.5 *1.2*	(e)
2.9 *0.58*	2.2 *0.54*	3.4 *0.88*	2.0	1.3 *1.14*	0.66 *0.90*	0.41 *10.7*						0.81 *0.77*	2.4 *0.86*	5.3 *0.90*	5.7 *1.0*	(f)
3.5 *0.72*	2.9 *0.71*	4.1 *1.1*	2.1	1.5 *1.28*	0.74 *1.0*	0.45 *12.0*	0.32 *9.9*	0.18 *4.1*	0.14 *2.3*	0.21 *6.4*	1.7 *1.4*	0.79 *0.70*	2.2 *0.80*	6.2 *1.1*	7.0 *1.3*	(g)

* Except as noted in footnotes 2 and 3.

† 1958.1 to 1967.2

‡ 1957.1 to 1967.2

§ Ratio of RMS to Mean Absolute Investment.

ables to changes in policy variables. In such a policy simulation we again take as given the four sets of variables described above; however, the time path of one or more policy variables is altered in some appropriate fashion. The response of any given endogenous variable to this "experimental" change in policy variables is then measured by the difference between the path of that variable in the dynamic simulation and in the policy simulation. In this paper we have chosen to focus attention on the response of GNP to changes in government expenditure and the channels through which this response is generated.

II. *The Multiplier Mechanism*

1. *Consumption Sector and the Derivation of Disposable Income.* As noted earlier, this mechanism arises from the interaction of two sets of relations: (a) the "consumption function" linking consumers' expenditure to their resources and (b) the set of mechanisms that links consumers' resources to GNP. With respect to (a), the distinguishing feature of our approach is that we treat total expenditure as resulting from both "consumption" and "investment" decisions. The first decision relates to the rate of consumption defined as expenditure on nondurables and ser-

vices plus the gross rental value of the stock of consumers' durables (all in constant prices). In line with our earlier work, we hypothesize that consumption is a linear function of (i) expected income net of taxes, approximated by a weighted average of real disposable income[4] current and lagged (up to twelve quarters) and (ii) net worth, also introduced with a short distributed lag (four quarters) to allow for a gradual response to capital gains.

The investment decision applies to expenditure on consumers' durables. Here we rely on a variant of the stock adjustment hypothesis in which the equilibrium stock is a function of consumption and the rental rate, but we also allow expenditure to vary in response to current saving (income minus consumption). The stock is derived recursively by subtracting depreciation and adding current expenditure. Finally, we obtain consumers' expenditure by subtracting from consumption the rental value of durables and adding back current purchases.[5]

How well our model fits the data can be judged from Table 1, columns 1, 2 and 3. The figures reported in row (*a*) are the RMS error resulting from a fifty-quarter dynamic simulation in which we take as exogenous disposable income, capital gains, the opportunity cost of capital (measured by the corporate bond rate), and prices, plus the initial net worth and stock of durables. The fit of the consumption function—column 1—is quite close. Investment in durables, like all investment flows, is harder to explain—column 2—but since it is a modest component of consumers' expenditure the error in the total—column 3—is again reasonably small, 0.6 percent. (This incidentally implies as error in personal saving, as defined in the national income accounts, of less than 10 percent, a fairly creditable performance for such a long, dynamic simulation.) Figure 1, Graph A, compares actual consumers' expenditure with the computed values.

To close the multiplier "system" we need to establish the relation of disposable income to GNP. This is accomplished by adding to and subtracting from GNP in current dollars (the product of GNP in 1958 dollars and its deflator) a number of items. These items consist essentially of taxes and transfers (the so-called "built-in" stabilizers) and corporate retention in addition to a few items taken as exogenous.[6] Federal government taxes and social security contributions and benefits are explained by equations containing (i) the relevant tax rates, (ii) some approximation to the appropriate tax base derived directly or indirectly from GNP or unemployment, and (iii) other legal and institutional arrangements. Corporate retention, as well as corporate taxes, depends on corporate gross profits. The derivation of the latter variable involves basically the estimation of total compensation (including the imputed compensation of

---

[4] We use total disposable income rather than disposable labor income, because of the difficulty of imputing personal income taxes between labor and non-labor income, particularly for the purposes of simulating different tax policies. However, our concept of disposable income is different from that of OBE because (1) we treat the federal personal income tax on a liabilities basis rather than on a payment basis and (2) it includes imputed gross rent on the stock of consumers' durables but excludes interest paid by consumers.

[5] Our consumption function and durable equation imply a unique relation between expenditure on nondurables and services and the arguments of the above two functions. In formulating and estimating our equations, we have taken this constraint into account (see [5]).

[6] Items taken as exogenous at the moment are: capital consumption allowance, business transfer payments, statistical discrepancy, subsidies less current surplus of federal (but not state and local) government enterprises, government transfer payments other than unemployment insurance benefits, and interest and rental income.

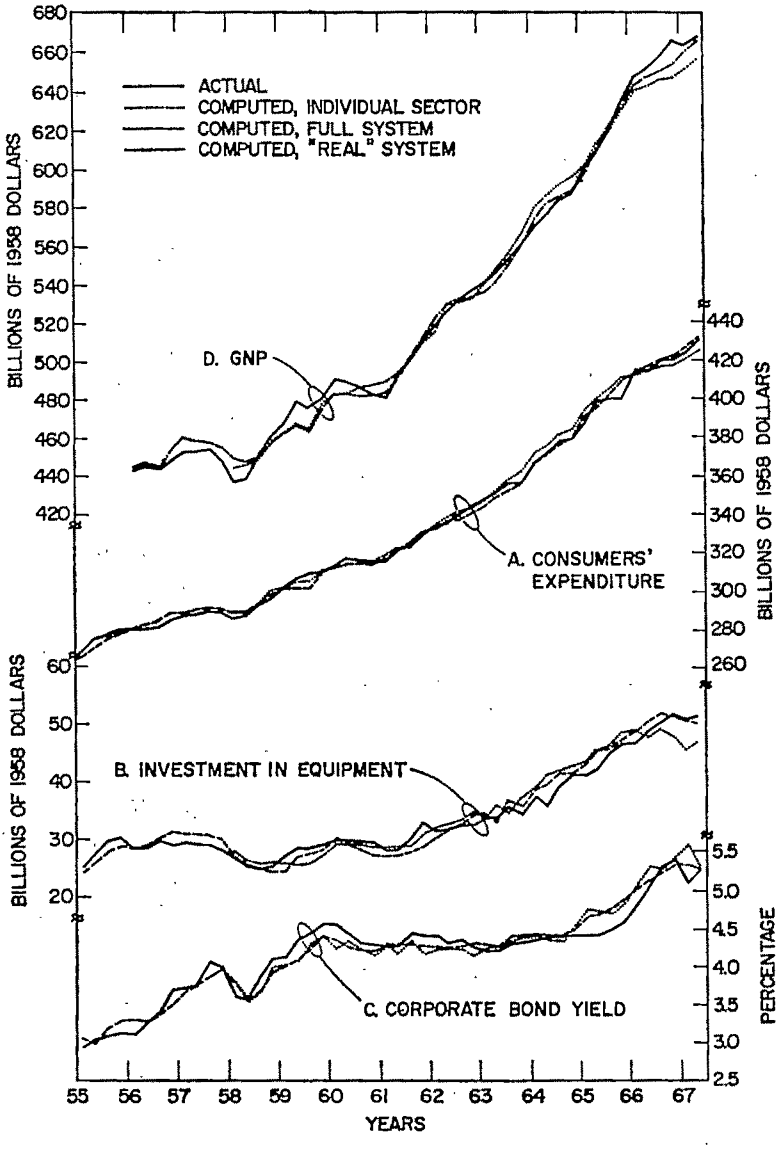


FIGURE 1

ACTUAL VALUES, SECTOR AND SYSTEM SIMULATIONS, SELECTED VARIABLES

self-employed) which is derived from an estimate of man-hours (cf. Section III.2.*b* below) and of hourly wages (cf. Section V). In row (*a*) of Table 1, columns 11, 12, and 13, we exhibit the RMS error of personal and disposable income computed by taking as exogenous GNP, the GNP deflator, the wage rate, in addition to the appropriate tax rates and the few minor items mentioned in footnote 6 above. Column 14 gives the error in our estimate of the federal government surplus or deficit on income and product account, which serves as a useful overall check on the performance of our federal taxes and transfers-to-persons equations, since the other federal government receipts and expenditures are exogenous. While the computation of a percentage figure is rather meaningless here since the base is a number around zero, the RMS of some one billion can be regarded as quite encouraging. The errors in man-hours, employment and unemployment reported in columns 15, 16, and 17 are also relevant since these variables play a role in the mechanism linking output (GNP) to disposable income. Unemployment is derived from employment and a labor supply equation which accounts for the labor force in terms of the size of population and its age and sex composition and a distributed lag of employment. The implied responsiveness of the labor force to employment accounts for the fact that the RMS in column 17 is appreciably smaller than that of column 16, though the relative RMS is much larger because of the smaller base. This relative error incidentally is also a close approximation to the error in the unemployment rate.

2. *The Pure Consumption Multiplier.* We can now examine the performance and implications of the consumption multiplier system, consisting of the consumption sector and the equations relating GNP to disposable income. It accounts for GNP, employment, consumption and the other variables mentioned in II.1 above in terms of the exogenous variables, the policy variables and three sets of experimentally exogenous variables, to wit, (i) all components of GNP other than consumers' expenditure, (ii) prices and wages, (iii) all financial variables. The RMS errors of a fifty-quarter dynamic simulation of this system are reported in row (*b*) of Table 1 for a number of important variables. As we had anticipated, the errors for each of the variables are larger than that reported in row (*a*). This larger error reflects both errors of the "independent variables" due to other equations of the system and "multiplier" effects. Thus, if, say, the disposable income subsector exaggerates the disposable income corresponding to a given GNP, this will not only create a direct error in disposable income but, in addition, it will cause an overestimate of consumption, which will cause a similar error in GNP raising again disposable income and so on; furthermore, the resulting final error in disposable income will tend to cause errors in consumption in several future periods, each of which is subject to a multiplier amplification.

In the light of the above considerations, it is rather reassuring that the system errors of row (*b*) are only moderately larger than those of row (*a*). The greatest amplification, about 30 percent, occurs in consumers' durables, presumably because this category is sensitive to the first difference in income, which is hard to account for with precision. On the other hand, the error of consumers' expenditure increases by only some 10 percent to just over two billion. Note that this number represents also the error of GNP, reported in column 26, since all other components of GNP are taken as exogenous.

With this background information we can now turn to an examination of the workings of the multiplier mechanism which is illustrated by curve *b* in Figure 3. To generate this curve, we first increased federal government expenditure on goods by $5 billion (in 1958 prices) in the first quarter of 1959, a point chosen in part on the ground that there was at the time a considerable amount of slack in the economy, the unemployment rate amounting to some 5.6 percent. We then compute for that quarter and the following ones the difference between the track of GNP generated by a dynamic simulation of the system of row (*b*), and the track resulting from a dynamic simulation of this system corresponding to the higher expenditure. The difference is then divided by $5 billion to obtain the multiplier measure, standardized per dollar of expenditure, which is exhibited in Figure 3. The most striking feature of our result is the modest size of the pure consumption multiplier: the impact multiplier is but 1.2, and by the end of the first year it still amounts to only 1.3. The effect continues to build up slowly, but it only reaches some 1.5 by the end of the second year and 1.6 by the end of the third.[7]

Very similar results were obtained when the experiment was repeated at other points of time. In other words, not surprisingly, the multiplier mechanism is fairly independent of initial conditions. The major explanation for the low multiplier seems to lie in the high degree of built-in stabilization provided by the federal tax structure. Indeed, we find that within the first quarter, the net tax take of the federal government rises by 2.3 billion, or some 40 percent of the increase in GNP. This very high proportion reflects in part the high share of corporate profits in the increase in GNP and the 50 percent corporate tax rate. By the end of the two years, the incremental tax take is about 35 percent of the increment in GNP, and by that time the increase in receipts of state and local government (0.4 billion) also contributes to hold down the multiplier (since state and local expenditure is taken as exogenous).

3. *The "Extended" Multiplier.* In addition to consumers' expenditure there are two components of final demand whose nature is more akin to that of consumption than that of investment. The first consists of expenditures and revenues of state and local government, which are treated as endogenous in our model, as explained in Gramlich [9]. The other is imports, which is at present explained by a fairly conventional equation involving income and capacity utilization. This equation is not too satisfactory and is soon to be replaced by a more disaggregated model covering both imports and exports, being elaborated by William Branson. RMS errors for these two sectors are reported in row (*a*) of Table 1, columns 5 and 6. However, both figures are somewhat distorted by exceptionally high errors in 1967. In the case of imports, this error is readily traceable to special circumstances relating to trade with Canada and to increased military expenditure, which are not taken into account in the present formulation. In the case of state and local government, the problem is a purely technical one and is being eliminated.

The effect of endogenizing state and local government and imports on our error measure of performance is rather negligible, as can be seen from row (*c*) of Table 1. Indeed, the GNP error, as well as several others, declines. This is partly due to the effect of offsetting errors, but in part it reflects the fact that endogenizing imports tends to reduce the multiplier. A simulation experiment of system (*c*) similar to that carried out for system (*b*) (not reported in Figure 3 to avoid clogging that figure) indicates that the multiplier is lower at every point of time, the difference ranging between 5 and 15 percent.

III. *The Investment Sector, the Accelerator Mechanism and the Multiplier Accelerator Interaction*

This sector has received a great deal of our attention, and incorporates a number of innovations, sketched out below.

1. *Inventory Investment.* Here we distinguish between commodities whose demands are satisfied from current production and/or from finished good inventories

---

[7] The continued rise of income thereafter, though at a much slower rate, reflects the role of wealth in the consumption function. The rise can be expected to continue until the increase in tax revenue reaches asymptotically the increase in expenditure.

and those produced to order with a significant processing time. For the first type of commodities, which is operationally identified with consumers' expenditure, we rely on the conventional "flexible accelerator." However, the current change in sales has a negative effect on inventory investment, since it initially drains inventories to a greater extent than it stimulates production. For the second type of commodities, defined to include capital equipment and, in principle, defense procurement, the change in inventories is generated by a distributed lag of current and past orders. The weights are positive for the more recent orders, which tend to produce increases in raw material and goods-in-process inventories; they are negative for the more distant orders which result in current deliveries and hence reduce inventories. We have also looked hard for evidence of financial effect—both cost of funds and variations in rationing of commercial loans—but so far with no success [10] [12].

On the whole, the model accounts reasonably well for a component of demand whose "mercurial" character is well known, though it makes some occasional very large error (e.g., 1966.3 and 4) which is reflected in the RMS in row (*a*) of Table 1, column 6.

2. *Investment in Plant and Equipment and Its Relation to Employment and Prices.*

a) *Basic Assumptions and Implications for Investment.*[8] It is a distinguishing feature of our model that investment, employment and the relation of prices to wages are derived consistently from the following four basic assumptions:

(1) *Neoclassical Production Function.* Output can be produced by a variety of technologies involving different factor intensities, and the set of such efficient technologies can be summarized by means of a neoclassic production function with constant returns to scale and constant elasticity of substitution, which shifts through time because of technical progress. (On the basis of empirical results, the constant elasticity was finally narrowed down to unity, implying a Cobb-Douglas function.)

(2) *Cost Minimization.* Additions to productive capacity at any point of time will occur through investment in that technology which minimizes cost per unit of output given the wage rate for labor and the rental rate for capital.

(3) *Ex post Nonsubstitution* (*Putty-Clay Model of Capital*). Once a specific technology satisfying the production function has been chosen reflecting the technical knowledge and relative prices prevailing at the time, the capital invested in that technology becomes irrevocably specialized and throughout its life can produce output only by being combined with labor in the initially chosen fixed proportion.

(4) *Fixed Markup on Long-run Minimum Average Cost.* Competition (entry) ties the price of output to long-run minimum average cost, but because of oligopolistic limitations on entry, the two need not be equal. Rather, price can exceed average cost by a factor which is assumed to be reasonably stable over the relevant run, though subject to short-run variations in response to variations in the rate of utilization of existing capacity.

Disregarding for the moment lags in decisions and execution, these assumptions imply the following relation among gross investment $I(t)$, the rental rate $R(t)$ and desired additions to capacity $\Delta\text{cap}(t)$

$$\iota(t) = \frac{A}{R(t)} \Delta \text{ cap } (t)$$

where A is a constant depending on the parameters of the production function and the mark up factor. We next hypothesize (i) that the desired (gross) addition to capacity at time t can be approximated by a distributed lag of past changes in output and of past levels of output, the latter reflecting replacement requirements; (ii) that some time is required for the chosen technology to respond to the rental rate, and therefore the current choice depends on a distributed lag of past values of R; and (iii) that a certain significant span of time is required to manufacture the capital goods once they have been ordered. This leads to two equations of the following type:

---

[8] For detailed discussions of this subject, see [3] [4].

$$O(t) = \sum_{\tau} a_{\tau} R^{-1}(t - \tau) \sum_{\tau} b_{\theta} X_{t}(t - \theta) \tag{1}$$

$$I(t) = \sum_{\tau} \omega_{t} O(t - \tau) \tag{2}$$

where $O(t)$ is orders placed, $I(t)$, expenditure on (deliveries of) capital goods, and X (t) output in quarter t. a and b are constants, such that $\Sigma a_T = A$, and the ω's are weights satisfying the constraints $\Sigma\omega_T = 1$.

The implications of our estimated new orders and investment equations are illustrated by means of three graphs. Figure 1B shows actual investment and its computed values, taking as exogenous the actual course of output (measured by real gross private domestic business product) and the rental rate. The RMS error is shown in row (a) of Table 1, column 8. The error for new orders, shown in column 7, is somewhat larger than for investment, which is an average of past orders.

Figure 2A depicts the response of new orders and investment to a ten billion step increase in output. Curve A shows that new orders respond promptly as the increase in sale generates a need for new capacity to accommodate the added output. They stay high for a while as the perceived need for new capacity responds only gradually to the increase in sales, but eventually settle back to a steady-state level which represents replacement requirements. The aggregate of new orders placed before the pure replacement regime amounts to 17 billion in our figure. This figure, as well as the constant rate thereafter, depends on the specific value of the rental rate. R—R is defined as the sum of the rate of depreciation and the required yield on capital times the price index of equipment relative to that of output. The required yield in turn is estimated along the Modigliani-Miller line [13], as a weighted average of corporate bond yields and stock yields. This basic formula is then adjusted to reflect relevant features of the tax system, including depreciation rules and investment credit.

Curve A is based on the rental rate R prevailing around 1960, which happens to be unusually high. If R were x percent lower, the whole graph would be x percent higher.

Curve B in Figure 2A shows the response of investment itself, which lags behind new orders because of processing time. Thus, in contrast to new orders, expenditure reaches its peak only about one-and-a-half years after the rise in sales, though, by construction, all orders eventually show up in expenditure. The shape of the expenditure curve can vary over time, because in our model the weights ω_τ of equation (2) depend on the rate of utilization of the equipment-producing industry as measured by the ratio of unfilled orders to sales. Our figures assume a rate of utilization somewhat below average. With a higher rate, the whole curve B would be pushed toward the right.

Figure 2B illustrates the effect on investment of a step change in the rental rate itself. Given the time path of desired additions to capacity, an x percent change

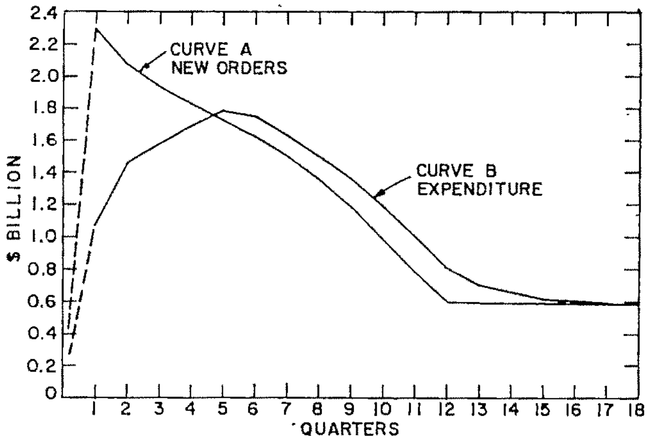


FIGURE 2A

RESPONSE OF EQUIPMENT TO $10 BILLION INCREASE (IN CONSTANT DOLLARS) IN PRIVATE GNP (LEVEL OF R AT 1960 VALUE)

in R eventually will change the level of investment permanently by precisely x percent. However, this reaction takes time and the figure shows what proportion of the final effect has become manifest t quarters after the change in R (assuming for convenience a constant rate of output and of desired capacity). The lags indicated by Figure 2B are quite long, and they bring out one of the important provisional findings of our research; namely, that several of the quantitatively important effects of financial variables controllable through monetary policy work with quite long (and variable) lags.

The response of new orders overshoots the steady-state level in the ninth and tenth quarters, but only very slightly. Since the standard neoclassical assumption of *ex post* as well as *ex ante* substitu-

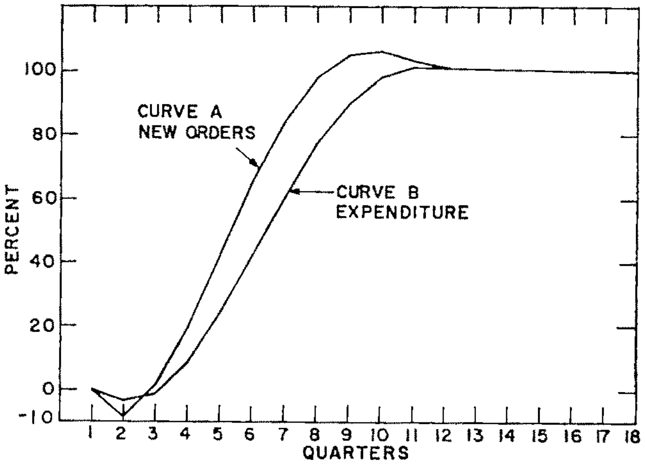


FIGURE 2B

RESPONSE OF EQUIPMENT TO A CHANGE IN RENTAL RATE: PERCENT OF STEADY-STATE EFFECT ATTAINED AFTER t QUARTERS WITH CONSTANT OUTPUT tion leads to an investment function with strong overshooting or "acceleration" effects in response to a change in R, the shape of the curves in Figure 2B can be regarded as significant evidence in support of the putty-clay model, not only in investment, but also in the short-run production function and pricing equation.

We may finally note that the elasticity of the rental rate R and hence of investment, with respect to a change in long-term interest rates, calculated in the neighborhood of prevailing long-term rates—say, around 4.5 percent—is in the order of 40 percent.

Investment in plant is treated pretty much along the lines of equipment. The empirical results imply that the response of plant investment to a change in R involves significantly greater overshooting than in the case of equipment, suggesting that plant is less specific and frozen in its use than equipment.

b) Implications for Employment (Short-run Production Function) and Prices. It follows directly from our postulates that, given the history of past investment, there is a well-defined direct labor requirement for any given level of output at any moment of time. In addition, however, there are indirect labor requirements to maintain the physical and managerial organization of the economy, which are more a function of the size of the productive capacity than of actual output. Since we cannot separate these two types of labor in our data, our employment function relates total man-hours to output, capacity, and the productivity trend. In order to incorporate the short-run dynamics associated with this process, we add to our equation the rate of change of output to reflect higher productivity of overtime hours and the unemployment rate to reflect labor hoarding.

These postulates, together with unitary elasticity of substitution in the production function, also imply that the price of output (net of indirect taxes) should be proportional to unit labor cost—the wage rate times man-hours per unit of output—associated with the technology that is currently most economical. Unit labor requirement, or its reciprocal, labor productivity, may be expected to change over time, due to technical change and, in principle, to variations in capital intensity induced by variations in the rental rate.

In the empirical equation we have assumed that the changing productivity can be approximated by a broken time trend and current productivity. In addition, we allow for the possibility that the adjustment of price to its equilibrium level may take more than one quarter, by using the lagged dependent variable. Finally, we hypothesize that when the ratio of demand to capacity becomes unusually high or low, a relaxation of the oligopolistic discipline will tend to give rise to an upward or downward shading of the markup. Empirically, we measure this demand pressure by the ratio of unfilled orders to shipments. Column 23 of Table 1 suggests that our formulation fits the historical facts quite well.

3. *Residential Construction.* The basic notion here is that most houses are built by builders for immediate sale for profit and, therefore, housing starts depend on the relation between the expected sales price of the house and construction costs, modified by availability of financing. The price of houses depends on the net rental rate of houses (gross rental less depreciation and taxes plus expected capital gains) and the appropriate discount rate. The gross rental rate depends in turn on the final demand for houses by consumers and the available stock of houses. The availability of financing is related to the inflow of funds into savings institutions. Unfortunately, the implementation of this approach proved to be extremely difficult,

in part because of lack of reliable data on price and rent of houses. As a result, we have been forced to eliminate the price of houses, and, in the case of single family houses, the rent, from our system and work with the reduced form equations.

On the whole, the values of the estimated coefficients appear reasonable and the equations explain the historical data for the period 1957–67 reasonably well, as shown in column 10 of Table 1. Yet, we have some qualms about the reliability of the results, and consider them as provisional.

4. *The "Real" System and the Accelerator-Multiplier Interaction.* Row (*d*) of Table 1 reports measures of performance for the "real" system as a whole, which essentially treats as experimentally exogenous only the financial variables and prices and wages. For most sectors the system error is not appreciably larger than in row (*a*), with the notable exception of equipment investment. Yet, the GNP error roughly doubles relative to row (*c*), though it remains modest. Figure 1, graph D, confirms that the computed value of GNP tracks rather well the actual course, at least in its broader movements, providing some ground for confidence in the usefulness of system (*d*) as a description of the working of the real sector of the economy.

We now turn again to Figure 3 where curve *d* illustrates the effect of five billion increase in government expenditure. Comparison with curve *b* indicates that the interaction of multiplier and accelerator results in a considerably larger expansion of GNP beginning with the second quarter. Also, as one might expect from the theoretical literature, the response of the system is now cyclical. What is perhaps surprising is that the expansion continues for three years, and the ensuing contraction is brief and of rather negligible magnitude.

As can be seen from the component part chart of Figure 4, the smooth, trend-like behavior of GNP is a result of fairly pronounced waves in several components of demand—notably inventory, equipment, and, to some extent, plant—which happen to offset each other to a large extent. It must also be remembered that the

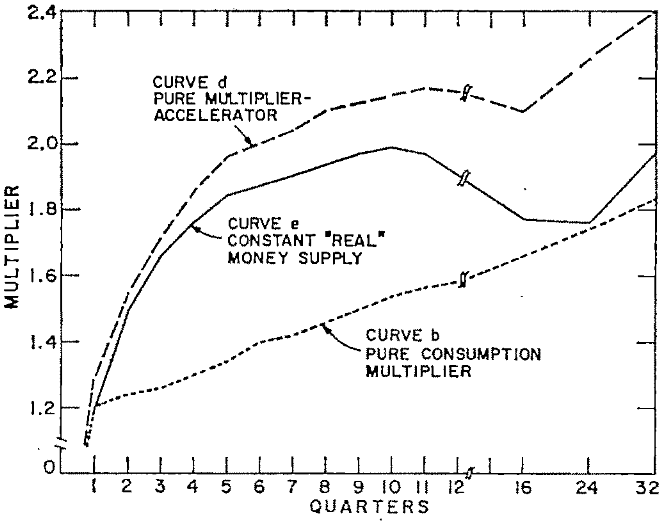


FIGURE 3

MULTIPLIER EFFECTS OF AN INCREASE IN GOVERNMENT EXPENDITURE

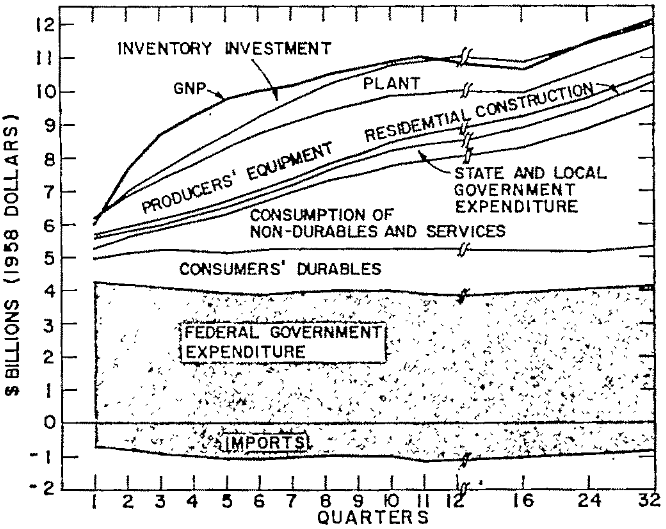


FIGURE 4

COMPONENT PARTS OF THE EFFECTS OF AN INCREASE OF $5 BILLION (IN 1958 DOLLARS) OF GOVERNMENT EXPENDITURE: PURE MULTIPLIER-ACCELERATOR MODEL

CALCUTTA UNIVERSITY CENTRAL LIBRARY

path is strongly affected by the response of government receipts to the increase in GNP which, under present fiscal arrangements, is remarkably high, implying a large "leakage." The marginal tax take is initially almost 40 percent and declines slowly to one-third; accordingly, even in the first quarter, almost half of the additional expenditure is recouped and the proportion gradually rises toward 70 percent. With a lower marginal tax rate, one would expect a more pronounced wave, though this remains to be established by appropriate simulation experiments.

IV. *The Financial Sector and Its Interaction with the "Real" Sector*

1. *Outline of the Sector.* The main link between the financial and the real sectors is provided by three long-term yields: the corporate bond yield, the mortgage rate, and the stock yield (approximated operationally by the dividend-price ratio). The stock yield in turn is explained (insofar as we succeed, which of course is not very far) by a distributed lag of the corporate bond yield, with a long-run elasticity close to one, the rate of change of the price level, and the rate of change of real dividends (as a crude approximation to expected capital gains). In the present version of the model the mortgage rate is also tied basically to the corporate bond rate (although work is almost completed which treats the mortgage market explicitly). The corporate bond rate is accounted for by a term structure equation in which the short-term rate is the input, as described by Modigliani and Sutch [15]. Finally, the short-term rate is determined by the equilibrium between the demand and supply in the money market. This market plays a central role, for it provides the mechanism through which the real sector feeds back on the financial markets contributing to the determination of market yields.

The working of the money market is described in a forthcoming paper [14]. Basically, the supply results from the behavior of commercial banks as they adjust their portfolio of earning assets to changing conditions, including changes in unborrowed reserves and in the demand for commercial loans.

On the demand side, the quantity demanded adjusts gradually to an equilibrium level which is determined, in the Fisherian tradition, by the level of transactions and the opportunity cost of holding money, measured by the short-term rate and the rate on saving deposits [2] [20] [11].

2. *Interaction of the Financial and the "Real" Sectors.* The performance of the financial sector is summarized in columns 18 to 22, row (*a*) of Table 1. Once again, the fit is generally rather good, with the understandable exception of the equity yield. We do not look for much improvement here, simply recognizing thereby the influence of hard-to-predict psychological variables on investment behavior, and thus, aggregate activity in the economy. Yet, if the stock market plays as important a role as our present results imply, then this suggests that other tools of policy may have to be relied upon to counteract possible unstabilizing behavior of this variable.

Row (*e*) of Table 1 reports errors of the system that result when the financial sector is joined to the real system covered in row (*d*). We now treat as exogenous unborrowed bank reserves and reserve requirements, the discount rate, and ceiling rates. The dividend yield is treated as endogenous insofar as it enters investment activity. However, capital gains on common stock are treated as exogenous (except those due to retained earnings of corporations) which is equivalent to taking the dividend yield as exogenous for the purpose of computing the market value of stock.

As might be expected, endogenizing the financial sector tends to increase the error of the investment equation. These errors, in turn, through the multiplier effects, raise somewhat the RMS of other equations and finally, the error of GNP to about 1.2 percent. On the whole, however, our model still tracks the historical path of the economy reasonably well.

Curve *e* in Figure 3 exhibits the effect on GNP of an increase in government expenditure implied by system (*e*). Comparison with curve *d* serves to indicate the extent to which an inelastic "real" money supply limits the effect of an expansionary fiscal action. We speak of real money supply, because we are still treating prices as exogenously given, which is roughly equivalent to assuming a monetary policy which accommodates changes in demand for money originating from rising prices, but not those arising from changes in real output.

From the figure the following implications of our model stand out: (*a*) the expansionary effects of the fiscal action are smaller throughout, although for the first three quarters or so, the differences are negligible; (*b*) even by the end of the second year, the difference is less than 10 percent; (*c*) eventually the difference does grow to be as much as 25 percent, but this takes a long time, in the order of five years. This long lag reflects the time required before a tightening of the short-term market is communicated to long-term rates, and the additional time needed before orders and finally investment respond to changes in the long rates.

V. *Wages, Prices, and Price Level Effects*

The relation of the price level to the wage rate has already been reviewed under III.2.*b*. The equation that explains the level of wages (the average rate of total compensation per hour of nonfarm private business sector) is in the tradition of Phillips, Perry, Simler and Tella [17] [16] [19]. It explains the rate of change of the wage rate by the rate of unemployment, the rate of change of corporate cash flow after taxes, and the rate of change of OASI compensation rate. We consider this specification much too crude and are currently considering a theoretically more satisfactory formulation, but so far we have not been able to improve on the present crude version, which performs surprisingly well (cf. Table 1, column 24).

We should point out one feature of our wage mechanism that distinguishes it from most other sectors. Because our wage hypothesis is formulated in the first differences, it does not imply an "equilibrium level." In other words, while for other variables we should expect errors to be transient and tend eventually to reverse themselves, errors in the wage rate could readily cumulate—except for the constraint arising from the nominal money supply and from a few variables which are not homogeneous of zero degree in prices, such as some real tax payments. But as should be clear from the results already presented in Figure 3, these financial effects are quite slow in their operation. Hence, when considering the entire system, including wages and prices, we cannot count on our simulation tracking the historical path closely for more than, say, two or three years; and, by the same token, little reliance can be placed on the result of policy simulations beyond that interval.

Empirically, it turns out that our wage-price sector tracks the actual course surprisingly well in a dynamic simulation running all the way from 1958 to the end of the period, even when the wage-price sector is joined to the rest of the system.[9] This can be seen from Table 1, rows (*f*) and (*g*). The first row shows the RMS resulting when the wage-price sector, but not the financial sector, is added to the "real" system (*d*). The RMS for price, though much higher than the single equation error, remains below 1 percent, and this is equally true of the wage error. The RMS of other sectors is not greatly affected, and that for GNP also remains below 1 percent. But, as one should expect, when the financial sectors are also endogenized, the presence of errors in prices does tend to magnify the errors of the individual real sector and of GNP. This deterioration can be traced largely to increased errors in those sectors which are most sensitive to financial yields—residential construction, plant, and equipment. These conclusions are supported by an examination of the entries in row (*g*) of Table 1. They are also confirmed by Figure 1, which shows for four critical variables the fit obtained in a dynamic simulation of the entire system, beginning in 1958.1.

Figure 5A exhibits the multiplier effects of government expenditure implied by the model with the wage-price sector endogenized. The effect is reported only for a span of ten quarters, since, as explained above, we do not feel that extrapolation beyond this range provides worthwhile information because the wage rate is endogenized.

Curve *d*, reproduced for reference from Figure 3, gives the multiplier path when both the wage-price and the financial sectors are taken as exogenous. Curve *f* is obtained by endogenizing prices and wages, but not the financial sector. Thus, a comparison of these two curves provides information on price level effects other than those operating through financial markets.

---

[9] In this simulation and all others in which prices are endogenous, we provisionally take as experimentally exogenous the ratio of unfilled orders to sales of "machinery and equipment" which enter the price equation (cf. III.2.*b*). This is because the relation of this variable to new orders for expenditure on equipment generated by the investment sector has not yet been worked out to our satisfaction. Note also that a simulation of the full system beginning in 1955.1 does tend to get off track because of cumulation of errors in the wage rate, though it eventually does get back on track by the early 1960's.

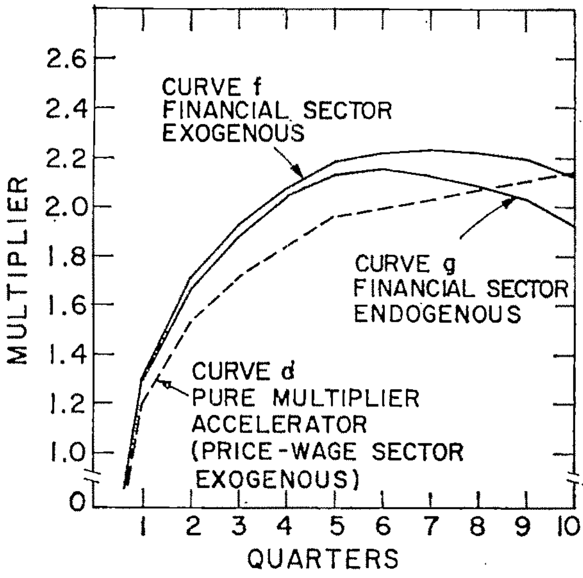


FIGURE 5A

MULTIPLIER EFFECTS OF AN INCREASE IN GOVERNMENT EXPENDITURE: PRICE-WAGE SECTOR ENDOGENOUS (WITHOUT CORRECTIONS FOR CAPITAL GAINS EFFECTS)

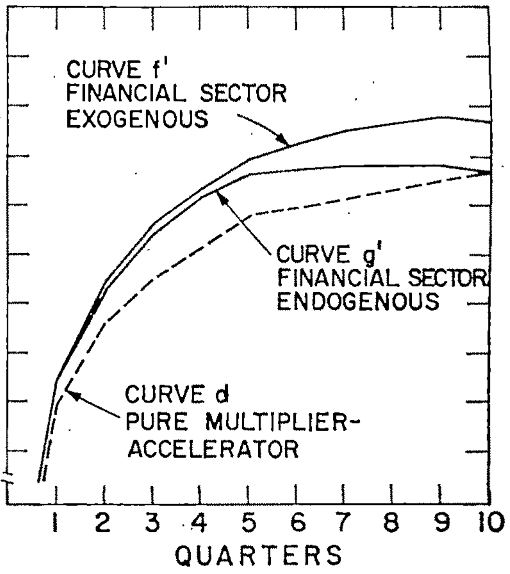


FIGURE 5B

MULTIPLIER EFFECTS OF AN INCREASE IN GOVERNMENT EXPENDITURE: PRICE-WAGE SECTOR ENDOGENOUS (WITH CORRECTIONS FOR CAPITAL GAINS EFFECTS)

It is apparent that endogenizing the price sector speeds up somewhat the response of GNP to the fiscal stimulus. This modest differential effect can be basically traced to the dependence of the wage rate on the corporate cash flow coupled with the lag in the adjustment of prices to changes in wages. These two features jointly have the short-run effect of increasing the responsiveness of labor income and, hence, personal and disposable income to the change in GNP. This effect should be transitory, and curves *d* and *f* do come to coincide by quarter 10.

The above conclusion must be qualified, however, because of a deficiency in the present specifications of the mechanism of adjustment of net worth to deviations of the price level from its historical path. Because in the simulations reported here we take as exogenous capital gains (except those attributable to corporate retention), the additional nominal capital gains resulting from a higher price level are not reflected in net worth. It is possible to make a crude adjustment for this shortcoming, and the multiplier path with this correction is exhibited by curve *f′* in Figure 5B. It is seen that the differential is longer lasting and begins to dwindle only by the ninth quarter.

Curve *g* in Figure 5A exhibits the response of GNP to the same increase in government expenditure when the financial sector is also endogenized, with the supply of unborrowed reserves kept at its historical level. Curve *g′* in Figure 5B exhibits the same multiplier path with an estimated correction for the net worth effect. As in the case of the system with prices exogenous, we should expect an inelastic money supply to counteract the fiscal stimulus, since higher prices along with higher output should increase the demand for money and hence, the level of interest rates, unless unborrowed reserves respond to the increase in demand for money. Furthermore, the structure of our model implies that, as long as real income and employment remain higher, wages and prices and, hence, finally the level of interest rates, will keep rising relative to the historical path. This in turn will exert an increasingly damping effect on those components of demand that are responsive to market yields, and will tend, in the end, to offset the expansionary effect of the increased expenditure.

The important question, however, is how long it takes for this offsetting action to take place and how far GNP can rise in the interim. This is precisely the question that Figure 5 is designed to answer in terms of our model, and the distance between curves *f* and *g* in Figures 5A and 5B can be thought of as a measure of the constraining effects resulting from the money supply. In both of these figures, this distance is rather negligible for at least a year, while the multiplier reaches a value of around 2. Thereafter the distance grows, but very gradually so that by the end of the tenth quarter it amounts to only about 10 percent of the size of the multiplier.

When both prices and the financial sector are made endogenous, the multiplier path implied by our model is no longer independent of initial conditions and the size of stimulus. In particular, the small difference between curves *g* and *f* noted above may in part be due to the relatively slack conditions of the economy in 1959 when the expenditure change is assumed to have taken place. We should expect a higher rate of utilization of resources to generate faster and larger increases in prices, producing in turn more pronounced effects on the cost and availability of funds. The higher utilization should also cause greater delays in deliveries of orders placed. For both of these reasons, the expansion of real output for a given fiscal stimulus should be smaller. Some preliminary experiments confirm these inferences, though they also suggest that the effects of differences in initial conditions are rather limited within the first year or so, becoming quite important thereafter.[10]

In concluding, we wish to stress again the tentative nature of all quantitative results reported in this paper, especially those involving the price-wage sector. In the first place, as noted in our discussion, the present version of the model contains several shortcomings, such as the treatment of imports, capital gains, and unfilled orders, and they are particularly troublesome in policy simulations including the price-wage sector. The work necessary to remedy these defects is under way, which should enable us to eliminate them from our model shortly. Second, before we can understand and evaluate the characteristics and the performance of the model, we must carry out more extensive simulations, both of a diagnostic type and of policy changes. In particular, while we have focused our attention here on the effects of a change in government expenditure, we need to analyze the implications of other policy changes, especially monetary policy.[11] We hope that this preliminary report will serve to shed light on the nature of the model and its potential uses as an analytical tool, and provide a basis for criticism and discussion.

---

[10] The results of these simulations are very preliminary and not worth reporting here especially since the ratio of unfilled orders to sales must be taken as exogenous for policy simulations at the moment. This shortcoming, which is being eliminated presently, is unlikely to distort our conclusions significantly if the experimental change in expenditure occurs in a relatively slack period, but can be serious under different conditions.

[11] For a preliminary report on the workings of monetary policy, see the parallel paper by DeLeeuw and Gramlich [21]. This model is also being used to analyze the effects of more complex policy changes such as a constant rate of growth of money supply.

REFERENCES

1. A. ANDO AND S. M. GOLDFELD, "An Econometric Model for Evaluating Stabilization Policies," in *Studies in Economic Stabilization*, ed. by A. Ando, E. C. Brown, and A. F. Friedlaender (Brookings Institution, 1968).
2. W. J. BAUMOL, "The Transactions Demand for Cash," *Q. J. E.*, Nov., 1952.
3. C. W. BISCHOFF, "A Study of Distributed Lags and Business Fixed Investment" (Ph.D. dissertation submitted to Massachusetts Inst. of Tech., Jan., 1968).
4. ———, *Lags in Fiscal and Monetary Impacts on Investment in Producers' Durable Equipment* (Cowles Found. Discussion Paper No. 250, June, 1968).
5. W. C. BRAINARD AND J. TOBIN, "Pitfalls in Financial Model Building," *A.E.R.*, May, 1968.
6. F. DELEEUW AND E. GRAMLICH, "The Federal Reserve-MIT Econometric Model," *Fed. Res. Bul.*, Jan., 1968.
7. G. FROMM AND P. TAUBMAN, *Policy Simulations with an Econometric Model* (Brookings Institution, 1968).
8. S. M. GOLDFELD, *Commercial Bank Behavior and Economic Activity* (North-Holland Pub. Co., 1966).
9. E. GRAMLICH, "State and Local Government Expenditure and Their Budget Constraints," *Int. Econ. Rev.* (forthcoming).
10. D. JAFFEE, "Credit Rationing and the Commercial Loan Market" (Ph.D. dissertation submitted to Massachusetts Inst. of Tech., May, 1968).
11. M. H. MILLER AND D. ORR, "A Model of the Demand for Money," *QJ.E.*, Aug. 1966.
12. F. MODIGLIANI AND D. JAFFEE, *Theory and Tests of Credit Rationing* (to be published).
13. F. MODIGLIANI AND M. H. MILLER, "Some Estimates of the Cost of Capital to the Electric Utility Industry," *A.E.R.*, June, 1966.
14. F. MODIGLIANI, R. RASCHE, AND J. P. COOPER, *Central Bank Policy, Money Supply, and the Short Term Rate of Interest* (to be published).
15. F. MODIGLIANI AND R. SUTCH, "Innovations in Interest Rate Policy," *A.E.R.*, May, 1966.
16. G. L. PERRY, *Unemployment, Money Wage Rates, and Inflation* (M.I.T. Press, 1966).
17. A. W. PHILLIPS, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," *Economica*, Nov., 1958.
18. R. H. RASCHE AND H. T. SHAPIRO, "The FRB-MIT Econometric Model: Its Special Features," *A.E.R.*, May, 1968.
19. N. J. SIMLER AND A. TELLA, "Labor Reserves and the Phillip Curve," *Rev. of Econ. and Statis.*, Feb., 1968.
20. J. TOBIN, "The Interest Elasticity of Transactions Demand for Cash," *Rev. of Econ. and Statis.*, Aug., 1956.
21. F. DE LEEUW AND E. GRAMLICH, "The Channels of Monetary Policy" (to be presented at the meetings of the American Finance Association in Dec., 1968).

DISCUSSION

RICHARD G. DAVIS: In "Money and Permanent Income: Some Empirical Tests," Tobin and Swan estimate an equation in which current real income is made a function of current prices and the current and one-period lagged money stock. The equation is derived from Friedman's hypotheses on the demand for money and permanent income and was first suggested in the well-known paper by Friedman and Schwartz. The statistical results produced by the Tobin-Swan estimation of this equation are indeed awful by every normal criterion. This is of course the authors' point. As far as I have been able to tell in the time I have had to study their paper, Tobin and Swan's procedures seem to follow the Friedman-Schwartz program faithfully, and their very negative verdict on the result seems wholly justified.

Later in their paper, Tobin and Swan use a variant of the Friedman-Schwartz equation to "predict" quarterly and annual changes in money income in the 1959–68 period. The results prove inferior even to "naïve" prediction techniques. In this part of the paper, however, the Tobin-Swan approach does not seem entirely fair. Instead of using parameters estimated from the "prediction" period or from the somewhat longer postwar period used in the earlier part of their paper, Tobin and Swan instead insert Friedman's own estimate of the permanent income elasticity of the demand for money (1.8) and his estimate of the weight of the current period's "measured" value in deriving the same period's "permanent" value of a variable (.33). This seems to put a very heavy burden on Friedman and Schwartz. Thus the Tobin-Swan prediction test is a test not only of Friedman and Schwartz's general hypothesis, but also of the applicability of the Friedman parameter values to the particular set of years covered by Tobin and Swan's tests. Their earlier results had already indicated that the Friedman parameters needed substantial adjustment for the postwar period. Moreover, the assumption implicit in the Friedman-Schwartz equation that the money market is cleared within each period may well be wrong as applied to quarterly periods, as Tobin and Swan themselves acknowledge, thereby weakening the value of the equation for quarterly predictions.

Even agreeing that the Tobin-Swan "prediction" tests may be rather unfair, it remains clear that the Friedman-Schwartz equation does work quite badly as applied to postwar data. Tobin and Swan introduce a new equation which implies dependence of the demand for money on current interest rates and expected price changes as well as on the permanent income and permanent price variables of the Friedman formulation. This revised equation, which Tobin and Swan characterize as "Keynesian," does much better than the Friedman-Schwartz equation in terms of the usual statistical criteria. Furthermore, it is interesting to see that when interest rates and price expectations are introduced, measured and permanent income turn out to be almost identical; that is, lagged measured values have only a small weight in the computation of current permanent values.

It is not, however, completely clear to me how Friedman and Schwartz would react to the apparently superior performance of the Keynesian version of their equation. These authors themselves suggested that their equation "should be expanded to include interest rates and perhaps the rate of change in prices," since these variables do have statistically significant effects on money demand and a fairly regular cyclical pattern and even though Friedman and Schwartz believe them to be much less important than permanent income ("Money and Business Cycles," p. 59). It is true, as Tobin and Swan point out, that once variables other than income are admitted into the money demand function, sources of fluctuations in income other than fluctuations in money must be acknowledged, at least as a theoretical possibility. Friedman has, however, agreed on numerous occasions that nonmonetary sources of disturbance are not only theoretically possible but undoubtedly of at least some practical importance. The substantive issue therefore remains whether or not money is the dominating exogenous influence on the cyclical course of business. While the Tobin-Swan paper contributes a negative vote to this question, their work is of course hardly likely to end the debate, tied as it is to a very particular version of the monetary view and indeed to particular parameter values as far as the prediction part of their paper is concerned.

The dispute about the importance of money has been growing more heated, especially in recent months when the economy has remained surprisingly strong in the wake of the tax surcharge. Tobin and Swan argue that "policy-makers and forecasters would not have much luck in trying to infer movements of money income from changes in money stock." Perhaps not, but they don't seem to have had much luck from anything else lately either. The fact seems to be that, despite the mass and variety of serious work bearing on the role of money in cycles now in existence, a decisive verdict on its importance relative to other factors is not yet available. A capsule survey of some of the unresolved problems will suggest why the matter still seems up in the air.

On the one hand we now have several large-scale econometric models which spell out a specific structure for the influence of money on business, chiefly through the effects of money (or reserves) on financial interest rates and on the availability of credit from banks and other intermediaries, and, in turn, the influence of these factors on specific components of aggregate demand. We also have numerous single-equation studies examining the role of interest rate and credit availability variables on the demand for the various components of current output. Over the years such studies have tended increasingly to find significant money supply effects, via interest rates and credit availability, as the statistical techniques employed and the lag structures considered have become more sophisticated. However, the interest rate effects uncovered by these studies appear generally to operate with a long lag. Moreover, contrary to the money supply school, the econometric models generally show fiscal parameters to be as powerful as, or more powerful than, monetary parameters in their effects on the economy.

In general, the proponents of the money supply view have been quite unwilling to accept the results of this type of research. It is certainly possible, if one so chooses, to become discouraged about the prospects for structural models. Thus their value for projections has so far seemed quite limited and there are of course many serious statistical problems. One complaint frequently voiced by the monetarists, including Friedman and Schwartz, against the standard econometric work, is an allegedly inadequate or nonexistent treatment of the implicit interest rates on consumer and producer durable goods in the demand equations estimated for the various GNP components. Their argument seems to be that while monetary policy does initially influence financial interest rates, it is the resulting changes in the spread between these rates and existing implicit interest rates on real capital that induce people to shift into or out of the latter. As long as implicit rates are left out of the econometric equations, parallel movements of these implicit rates with financial interest rates may be masking a much sharper and quicker reaction of demand to changes in relative interest rates (i.e., financial versus implicit) than equations containing only financial rates seem to show. This doesn't sound too likely to me, but it is certainly not possible to settle the matter on the basis of existing research.

In opposition to the econometric literature, the money supply proponents have offered N.B.E.R.-type cyclical analysis, as exemplified by the work of Friedman and Schwartz and of Cagan, and what has been called the "reduced form" approach as exemplified by Friedman and Meiselman and by recent work done at the Federal Reserve Bank of St. Louis by Anderson and Jordan. The Friedman-type cyclical analysis has never seemed to me to grapple adequately with the implications of the existence of a strong reverse effect running from business to money. The existence of such a reverse effect undercuts seriously the evidential value of the historical timing pattern of business and monetary cycles in establishing the causative role of money. While Professor Cagan evidently believes that his work on the determinants of the money supply substantially supports the Friedman position, my own view is that its implications are, in fact, rather subversive. Cagan's work suggests rather clearly that the characteristic cyclical timing relationships between monetary rates of change and the business cycle are very importantly conditioned by the influence of business on money. It seems to me that the case is made even stronger if one takes explicit account of the impact of the Federal Reserve's attempts at countercyclical policy, especially in view of the System's tendency to concentrate on money market variables as short-term operating targets. Federal Reserve behavior alone may be sufficient to explain the characteristic lead of cycles in the monetary growth rate ahead of business cycle turning points during the postwar period without the need to posit any influence of money on business whatever.

In the spirit of the Friedman-Meiselman "reduced form" approach is the Anderson-Jordan work mentioned a moment ago. These authors have estimated quarterly changes in GNP as a function of current and one- to three-quarter lagged values of changes in the narrowly-defined money supply and the full employment surplus. They find little role for the fiscal variables, but an important and rather quick-acting influence of the money supply. I doubt that their dismissal of fiscal policy should be taken too seriously, but I find of some interest the ability of changes in the money supply, current and lagged from one to three quarters, to explain postwar quarterly changes in GNP reasonably well. For the 1952–68 period, the adjusted R^2 is about 0.5, roughly twice that obtained for quarterly changes in the Tobin-Swan estimation of the Friedman-Schwartz equation. Just for fun, I tried to reverse the direction of causation implied by the St. Louis formulation; that is, to explain current changes in money in terms of current and lagged changes in GNP. The results are clearly inferior to the case where money is assumed to be the driving force. Moreover, in the equations I tried, the lagged values of changes in GNP contributed little or nothing to explain current changes in money.

I do think the results turned up by the economists at the St. Louis Fed deserve to be contended with. Is it in fact true that this quarter's change in GNP is very heavily influenced by the concurrent and previous quarter's change in the money supply, and, to a lesser extent, by the change in the two quarters before that? Such a conclusion, with its implication of very short lags, seems to fly in the face of all we thought we had learned from the econometric literature mentioned earlier and from surveys and interviews of the relatively limited sensitivity of businessmen and consumers in the short run to changes in interest rates. Indeed this conventional view was apparently supported even by the experience of the famous "credit crunch" of 1966. Surveys of its impact on business and on state and local spending decisions did find an impact from sharply higher interest rates and stricter rationing of bank loans. Impressionistically speaking, however, the impact does not seem to have been of anything like the order of magnitude that the St. Louis equations appear to imply.

And so the argument goes, back and forth like a ping pong ball. My own view is that the burden of proof is still on those seeking to challenge received doctrine. Reduced forms such as the St. Louis equations need to be examined very carefully for stability and predictive power. More important, proponents of the money supply view should certainly try to set out and estimate reasonably detailed structural models of their own to show how they would do it differently, especially as regards the treatment of implicit interest rates, possible wealth effects, and lag structures. In the meantime, however, I suppose practitioners of the not-so-fine art of economic forecasting can try to be eclectic. Perhaps they should prepare one set of forecasts by conventional procedures and a second set by means of some reduced form money supply equation—though not the one presented by Tobin and Swan! The final forecast should be a weighted average, with weights reflecting the forecaster's subjective probabilities on the final outcome of the debate currently in progress.

DAVID MEISELMAN: If I didn't know him better, I might suspect that our Chairman, Jack Gurley, enjoyed what Karl Brunner has called the battle of the radio stations, that is, between AM and FM, because Jack has asked that I reciprocate AM's critique of some of my work by commenting on their paper this morning rather than direct my remarks to the other two papers on the program. For those of you who may have been withholding judgment about who won the last encounter, I hope you don't mind if I emphasize who appeared at today's session and who didn't.

I regret that I won't be able to say much this morning, first, because neither Franco nor Albert are here, and, second, because I never did get the complete set of equations to work from. I do have their text, but when I really tried to track down many of the complex interrelations, I found that, in many respects, I was not able to do so in a way that satisfied me. I think that I understand many aspects of the model, but because the model is so complex and the text so brief, I think it best if I restrict my comments to less than I would wish.

Although the FRB-MIT research effort does not conform to my own research preferences and although I am generally skeptical about the fruitfulness of this general research strategy, I believe this research is an imposing effort by extremely gifted, devoted, and tenacious scholars that both ought to be done—thank goodness by somebody besides myself—and ought to demand the friendly and critical attention of a large audience of economists and others interested in the progress of economic science. For especially after recent advances in research technology, we shall never know whether a task as ambitious as this can succeed unless we try, and we shall never know what it takes to succeed unless we keep testing and improving.

In the old battle of the radio stations there was an important difference between the research strategies of AM and FM which we should keep in mind in evaluating this and similar work. The one research strategy is to construct and test a complex and sophisticated model employing many variables—the route Ando and Modigliani and their coresearchers have taken. The other research strategy is to construct and test relatively simple models using few variables—and note that simple is not to be confused with simple minded or simplistic—which can then be tested more readily in a wider variety of contexts. In other words, there are two general and in many respects alternative ways to allocate scarce research resources. Friedman and I opted for one and Ando and Modigliani and their coworkers have opted for another.

In both routes, hypotheses are to be accepted, rejected, or modified on the basis of new evidence or as better ways of viewing the evidence are developed. And, along both routes there may be important by-products of the research process. For example, important by-products of our work involved questions about selecting measures that are the appropriate empirical counterparts of commonly used theoretical constructs, such as autonomous and induced expenditures. Similarly, I suspect that the Ando-Modigliani and FRB effort has uncovered other kinds of problems and solutions, especially of a general equilibrium nature, overlooked in examining simpler models. I certainly hope that the researchers involved in this massive effort will take time away from their numbers to develop for the literature some of the general equilibrium relations, or hints of them, they have uncovered; in other words, that the research reports cover more than the results of a vast curve-fitting exercise.

For evaluating the results of the two research strategies, I also hope that the results of the elaborate 200 plus equation model are eventually compared with the results of simpler models. Some comparisons are obviously impossible because simple models tend to focus on answering one question, for example, something correspondingly "simple" such as predicting aggregate income, whereas the elaborate models try to answer many questions simultaneously, one of their great appeals and snares. Yet other comparisons are possible. We can get the two different animals to run the same race in terms of predicting aggregate income or perhaps prices, and similarly we can ask whether the elaborate models do better than naïve models that depend on extrapolating trends or some other autoregressive scheme.

More specifically, even without asking whether the differences in results are worth the costs involved, I wonder how much better this model has done in predicting aggregate income than some of the single equation monetary models. I have no way of knowing until the race is run, and although I may be dead wrong, my hunch is that the race would turn out to be a close one, with some likelihood that the elaborate model would get lost around a turn—a business cycle or interest rate turn, that is.

In the very limited time available for discussing substantive matters and with little to work from, there are only a few issues I can touch. First, I find that one of the most interesting and potentially most useful innovations is the welcome treatment of household expenditures, usually labeled consumption or consumers' expenditures, as resulting from both "consumption" and "investment" decisions. The "consumption" component is taken as expenditures for nondurables and services plus the gross rental value of the stock of consumers' durables, which is an approximation of the value of the stream of consumption services. These consumption expenditures are a function of expected real income plus net worth. Regarding net worth, I am puzzled by the treatment of capital gains and losses. Some capital gains and losses result from changes in interest rates. But there can be substantial capital gains and losses when interest rates or rates of discount vary hardly at all because of changes in expected equity returns. For example, in 1962 we observed substantial changes in stock prices when interest rates were relatively stable, which suggests that the price of equities in that year fell because of a downward revision of expected future earnings. Neglecting these capital gains would seem to be a serious omission.

In this model, as I understand it, one of the connections between interest rates and aggregate expenditures is through the capital gains and losses that result from changes in long-term rates impinging on consumption expenditures. There is another problem here in addition to the equity capital gains and losses I have just mentioned because the model seems to overlook the fact that households (consumers) have long-term liabilities as well long-term assets. The present value of these long-term liabilities also may vary with interest rates, so that net worth may be affected, but with a sign opposite to that resulting from changes in the present value of household long-term assets. In other words, it is necessary to look at both sides of the balance sheet to evaluate the consequences for net worth of a change in interest rates.

I also have some reservations about the fact that there are neither short-term interest rates in the demand for consumer durables nor monetary variables nor credit market variables either, unless the rental rate, which incorporates a long-term interest rate, is supposed to capture the effects of interest rates and monetary and credit variables. Ando and Modigliani report that the short-term interest rates do not seem to have any explanatory power. I would urge them to keep trying.

Although I am bemused by the Ando-Modigliani tentative conclusions that the government expenditure and other multipliers are quite low (making fiscal policy effects of government action weak) and that there are long and variable lags in the effects of monetary policy, I am disturbed by several other features of the model. First, I have some reservations about the Ando-Modigliani view of the banking system and the money supply mechanism, especially that borrowed reserves are endogenous, not a Federal Reserve policed or controlled operation. Thus, I may not accept their interpretation of what constitutes an exogenous monetary disturbance or a disturbance "caused" by Federal Reserve policy. Second, I take exception to the Ando-Modigliani view about the channels through which money and monetary policy work—rather, their presumption of fact that theirs is the only channel or temporal sequence for money and for monetary policy to affect other variables in the economy. This clearly biases many results.[1] As I understand it, a brief summary of the main "linkages" would go something like this: the supply and demand for money determines short-term interest rates; the change in short-term interest rates affects the long-term interest rate through a term structure relationship, but very slowly; the long-term interest rate affects aggregate expenditures through three long-term rates: (1) the corporate bond yield, where the corporate bond yields are nominal market interest rates, (2) the mortgage rate, which is tied to the bond yield, and (3) a stock yield, which is the dividend price ratio. The stock yield is explained by a distributed lag of corporate bond yields. The three yields thus essentially collapse to one: the corporate bond yield.

---

[1] See D. Meiselman, "Strotz and Minsky on Monetary Variables and Aggregate Demand," in *Monetary Process and Policy: A Symposium*, ed., George Horwich, 1967.

Long-term rates are understood to affect investment expenditures, including housing, and also to affect consumption through its effects on capital gains. Taken together, long-term rates influence spending and aggregate income which in turn affect the demand for money by altering transactions demands for cash. A change in transactions demands affects the short-term rate. The change in short-term rates in turn alters the long-term rate, etc., as the initial disturbance brought about by the change in the quantity of money tends to be systematically damped.

In addition, investment expenditures are also a function of what they term the "required yield": a weighted average of corporate bond yields and stock yields. I find this rather puzzling, especially in the context of trying to explain investment expenditures during a period of anticipated inflation, because anticipated inflation clearly affects bond yields but not pure equity yields. (Stock yields more nearly approximate real rates of return of debt-free corporations whose earnings and assets move with the price level.) In general, the sequence seems to be a very traditional Keynesian one, to which I can make the by now very traditional comments that money can be spent for all kinds of assets and services including goods and services directly, not merely for bonds, and here, only short-term claims at that. Of course, it is not necessary that the money be exchanged only for debt instruments. In addition, I would be very interested to see how this model would reconcile its results with the growing evidence that the demand for consumer durables is affected by monetary policy.[2]

I wish to raise some questions about a major problem in this and other macromodels, which is, how prices are determined as well as a few aspects of collateral questions regarding the role of price expectations. In common with many other aggregative models, as I understand it, the prices of output or the price level is taken to be essentially a fixed percent markup over unit labor costs (net of indirect taxes), here modified somewhat by a capacity-demand relationship. In the main, however, the price level is a function of unit labor costs, which, as I have pointed out elsewhere, seems to be a labor theory of the value of money. Unlike most other attempts to explain the price level, the Ando-Modigliani model has money in the system. Of course, this is a convenient way of joining Cobb-Douglas production function assumptions regarding the constancy of labor and capital shares in national income with a markup theory of prices. Note in this context that they assume, as I understand it, not measure, that there is a unit elastic demand for investment with respect to the rental rate.

It seems to me that, even if we accept the Cobb-Douglas production function as an accurate portrayal of equilibrium production relationships, this may not be an accurate portrayal of short-run disequilibrium situations where wages and prices have not fully adjusted, which, combined with the lack of full steady-state adjustment and perfect competition norms, may leave too much slack for useful short-run analysis, especially in this quarterly model.

I am curious what happens to consumption expenditures and investment expenditures during inflation in the model. Recall that capital gains are a variable in consumption decisions, and interest rates and price dividend ratios are relevant in investment decisions. To see some of the problems posed by inflation, consider first an unanticipated inflation generated by an excessive increase in the stock of money. Following the increase in the stock of money, short-term interest rates fall, then long-term interest rates fall, consumption and investment rise, aggregrate nominal income and expenditures rise which generates an increase in transactions demand for cash so that short-term interest rates rise, then long-term interest rates rise, too, tending to shut off some of the rise in consumption and investment, and all with very long lags. (The very long lags here hardly square with recent monetary experience.) During the process we may or may not overshoot with respect to short-term interest rates. I find this hard to accept, especially given what seem to be faster bond yield responses to changes in prices and changes in real output.

---

[2] For example, see Thomas Simpson, "Money, Portfolio Adjustment and the Demand for Consumer Durables" (unpublished, Univ. of Chicago, 1968).

Consider the second hypothetical case where we have a steady-state inflation which is fully anticipated. Under these circumstances we would expect that long-term interest rates would incorporate the expected inflation and be permanently "high" but not necessarily rising. Short-term interest rates would also be permanently high. However, in the Ando-Modigliani model, the long-term rate of interest may never really catch up. And, more important, the model predicts that investment expenditures would fall because nominal interest rates rise whereas price earnings ratios do not.

It seems to me that during anticipated inflation price earnings ratios are constant, but nominal interest rates rise by the anticipated inflation rate. Frictions and the like aside, there is surely no disincentive to invest, and no real capital gains in the aggregate.

On the other hand, as I have argued elsewhere[3], if we have less than full employment and sticky wage rates, an unanticipated increase in aggregate demand stemming from an increase in the stock of money leads to a corresponding unanticipated increase in prices which will generally be associated with a rise in both real and nominal rates of return as well as rise in capital outlays. The main mechanism in this interaction is the increase in employment, thereby in the ratio of labor to capital, and thereby in the marginal product of capital. This is equivalent to arguing that Hicks's *IS* curve has a positive slope.

Ando and Modigliani admit that they have capital gains problems and that the wage-price sectors are poor. Lest we be too hard on Ando and Modigliani, it seems to me that other models do not do very well predicting the price level either, or in allocating a predicted change in nominal income into the change in prices on the one hand and the change in real output on the other. I believe we all share the common temptation to treat all nominal values as if they were real values.

Regarding improvements that Ando and Modigliani may make in these areas, I look forward to modifications in their treatment of capital gains, the determination of prices, and changes in the wage-price sectors. In trying to improve these sectors I hope that they draw on the work of Phelps[4] and of Lucas and Rapping[5] which give more important roles to price expectations in wage determination as well as a better sense of the dynamics of the labor market than the paradigms they use.

I know you all join me in wishing them well on their econometric intergalactic flight. It seems to me that they still have a long way to go.

DONALD P. TUCKER: The paper by Cagan and Gandolfi raises several important issues of monetary economics, but I shall restrict my comments to one main point in order to be able to treat it in detail. The purpose of their paper is to present more evidence on the important issue of the length of the lag between monetary changes and the resulting income and employment changes that follow, but I shall present an argument that the inference they draw about income and employment lags does not represent valid economic reasoning. Their conclusion simply does not follow from their empirical results.

Their empirical results provide them with estimates of the coefficients of an equation that relates the interest rate to changes in the money stock, with the number of lagged money terms extending to 38. From this relation they find that an increase in the rate of monetary expansion causes the interest rate to respond as shown in their Chart I. It lowers the interest rate at first but then raises it back up above its former level after several months. It seems to take about six or seven months for the interest rate to reach the bottom of its decline before turning up again.

---

[3] D. Meiselman, "Money, Factor Proportions and the Real Cycle," presented at the Zurich Meetings of the Econometric Society, 1964.

[4] See Edmund S. Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium," *J.P.E.*, Part II, July/Aug., 1968.

[5] Robert E. Lucas, Jr., and Leonard A. Rapping, "Real Wages, Employment, and the Price Level: An Integrated Theory of Aggregate Supply" (unpublished paper, Carnegie-Mellon Univ., Mar., 1968), and "Price Expectations and the Phillips Curve" (unpublished paper, Carnegie-Mellon Univ., Oct., 1968).

So far so good. But this is not their primary concern. Their next step is to draw an inference about income and employment adjustment lags from their interest rate results. "The ratio of money to nominal income rises and produces a decline in interest rates *until nominal income begins to rise*. . . . This change in the direction of interest rate movements following a monetary disturbance reflects, and therefore provides a measure of, the lag in monetary effects on income" (italics mine). Since it takes about six months for interest rates to reach bottom after an increase in the rate of monetary expansion, it must take roughly six months before income begins to respond. This is their stated conclusion.

This is not a valid deduction, however, because short-run economic behavior is not necessarily that simple. Their conclusion is based implicitly on what they believe to be a property of the demand for money. They seem to believe that the demand for money will be satisfied only if the interest rate begins to rise as soon as the rate of growth of nominal income begins to rise. To demonstrate the weakness of their argument I am going to present a theoretical counterexample. This counterexample is a hypothetical but plausible money demand function having the property that the interest rate takes several months to hit bottom even when income responds very rapidly to the monetary stimulus. The turning point of the interest rate tells nothing about the length of time that must elapse before the rate of income growth responds.

The money demand function I have in mind has the form

$$M^d = d + eY^p - f(Y - Y^p) - gR^p - h(R - R^p), \tag{1}$$

where Y and R are current real income and interest rate, and Y^p and R^p are "permanent" real income and interest rate. This counterexample is meant to apply to short-run adjustments in a situation of less than full employment in which the price level is stable. The third term represents the monetary counterpart to the accelerator effect on investment. An increase in real income provides a stimulus to investment—an opportunity to employ funds profitably. Hence, the opportunity cost of holding idle cash is increased and money demand falls. The fifth term represents a speculative influence. When the interest rate falls, people expect it to rise again, and so they try to increase their money holdings in anticipation of a fall in bond prices. Y^p and R^p are weighted averages of past values of Y and R, with the weights first rising and then falling. The heaviest weight is given to values two or three months in the past, and these meaures have the property that

$$(1 - \lambda L)(1 - \theta L)Y^p = (1 - \lambda)(1 - \theta)Y \tag{2}$$

$$(1 - \lambda L)(1 - \theta L)R^p = (1 - \lambda)(1 - \theta)R, \tag{3}$$

where L is the lag operator, and where $0<\theta<\lambda<1$. In this formulation each weight is the difference between two declining exponentials, and this makes it second in the family of weighted averages that begins with exponentially declining weights.

Eliminating Y^p and R^p from equation (1) by using (2) and (3), and using $\dot{Y}$ and $\dot{M}$ for the first differences of Y and M, we get the third order difference equation

$$\begin{aligned} (1 - \lambda L)(1 - \theta L)\dot{M} &= [(e + f)(1 - \lambda)(1 - \theta) \\ &\quad - f(1 - \lambda L)(1 - \theta L)]\dot{Y} \\ &\quad - [(g - h)(1 - \lambda)(1 - \theta) \\ &\quad + h(1 - \lambda L)(1 - \theta L)](1 - L)R. \end{aligned} \tag{4}$$

TABLE 1

ADJUSTMENT OF INTEREST RATE, GIVEN A UNIT INCREASE IN THE RATE OF MONETARY EXPANSION AND AN ARBITRARY PATH FOR THE RATE OF EXPANSION OF REAL INCOME

(Figures for R are deviations from initial value)

Time	$\dot{M}$	$\dot{Y}$	R	Time	$\dot{M}$	$\dot{Y}$	R
0	0.0	0.0	0.00	12	1.0	5.0	−1.04
1	1.0	0.3	−0.31	13	1.0	5.0	−0.95
2	1.0	1.5	−0.61	14	1.0	5.0	−0.85
3	1.0	3.2	−0.89	15	1.0	5.0	−0.75
4	1.0	4.2	−1.10	16	1.0	5.0	−0.64
5	1.0	4.7	−1.25	17	1.0	5.0	−0.54
6	1.0	5.0	−1.33	18	1.0	5.0	−0.43
7	1.0	5.0	−1.35	19	1.0	5.0	−0.33
8	1.0	5.0	−1.32	20	1.0	5.0	−0.22
9	1.0	5.0	−1.27	21	1.0	5.0	−0.11
10	1.0	5.0	−1.20	22	1.0	5.0	+0.00
11	1.0	5.0	−1.12				

For my numerical example I have chosen the following parameter values: $e=0.33$, $f=0.1$, $g=5$, $h=3$, $\lambda=0.8$, and $\theta=0.4$. The values for e and g are close to the values I used in my 1966 paper on money demand lags ("Dynamic Income Adjustment to Money-Supply Changes," *A.E.R.*, June, 1966), and the other values are hypothetical. I have supposed also that $\dot{Y}$ has the behavior shown in Table 1 when the rate of monetary expansion $\dot{M}$ increases by one unit from zero to one and stays at its new value. Equation (4) may then be used to show that the deviations of the interest rate from its initial value will follow the pattern shown in Table 1.

The gist of the results is this. The response of the interest rate follows closely the pattern shown in Cagan and Gandolfi's Chart I—it declines for seven periods before reversing direction, and then it rises above its former level—but this occurs in spite of the fact that income responds strongly in the second period and reaches its maximum rate of growth in the sixth period. In fact, the rapid increase of $\dot{Y}$ in the second, third, and fourth periods is partly responsible for the fact that R continues to decline for so many periods. As $\dot{Y}$ rises, the third term in the money demand function (1) becomes stronger in the direction of reducing money demand. Therefore, the interest rate must continue to drop in successive periods to counterbalance this increasingly strong effect. The speculative fifth term in (1) also is important because it means that only moderate interest rate reductions are needed to boost money demand by a given amount; because the declines are moderate and the pattern of weights in R^p is humped, R^p does not build up much strength during the first few periods. Eventually, of course, Y^p and R^p begin to have more influence, but the behavior of R in the first few periods depends heavily on the third and fifth terms in (1).

I do not claim that the demand for money necessarily has this form and these parameter values in the real world. My point is not an empirical one but rather the theoretical one that the real world could look like this. Cagan and Gandolfi have not ruled out this possibility, and they can not rule it out by casual argument. They can rule it out only by an exhaustive empirical study of the lags in the demand for money, and in the absence of such a study the form of these lags remains unknown, and the possibility I have suggested, plus many others that could produce similar results, remains open. As long as these possibilities remain open, Cagan and Gandolfi's conclusion is not valid. The interest rate adjustment path they find tells us nothing except the adjustment path of interest rates.

THE FUTURE OF GOLD

REAL GOLD, DOLLARS, AND PAPER GOLD

By ROBERT A. MUNDELL
University of Chicago

The alternatives implied in my title lead to different worlds and we have to decide which is better. The issues are subtle and involved and I can do no more than sketch answers to them. I shall argue that: (1) The accelerating frequency of exchange crises in the system over the past eight years is due to the undervaluation of gold. (2) The so-called U.S. "balance-of-payments deficit" is the result of the growth in world demand for reserve media which, since 1958, could only be supplied by the U.S. and is not the consequence of a faulty adjustment mechanism. (3) Elimination of the system of crises could be accomplished either by an increase in the real price of gold or by the introduction of paper gold; but the simultaneous elimination of the sequences of crises and the U.S. balance-of-payments deficit can only be resolved by the introduction of paper gold. Therefore a solution through paper gold is better than a return to real gold.

I. *The Cause of Crises*

The system of periodic crises that began in October, 1960, and continued through 1968 is not over. We should anticipate further major disturbances in 1969. Crises have become permanent stage props of the present system. An international crisis is one means by which a government can dramatize the need for an alteration in its policies and shift much of the blame for unpopular features of the new policy from the national government to the international community or to the evil little goblins dedicated to the lust for gold.

It would be hard for a government to create a crisis out of thin air. But if the groundwork has been properly laid, a government can create a crisis in about two months without showing unseemly haste. In the summer of 1968, for example, there was a steady outflow of capital from France stimulated by the May-June uprisings and the wage settlements. But the signal for the crisis of November was the lifting of exchange controls in early September combined with the massive credit expansion inside France. The French worker had to be punished for his bad behavior in May and June and the November crisis was the best way of making the punishment suit the crime.

I am being partly facetious, of course, about the deliberate engineering of crises. It would be, perhaps, more judicious to say that the system as a whole has become less tolerant of mistakes in policy. The power in the hands of private financiers has risen with the massive increase in international floating capital over the past few years. As long as the national reserves of each of the major countries were large relative to this private floating capital, a single country in isolation could withstand an attack. But the reserves of a large number of countries in the system are small in relation to the capital outflow pressure that can build up even in a few weeks. To keep control of the system now requires the collective reserves of the major countries. The central bankers who meet once a month in Basle have really turned themselves into something like a floating first aid station, rushing monetary supplies to the next victim of speculative opinion. The actual international reserves held by the major countries are small now but the velocity of these reserves has been increased. We have fewer reserves in relation to the demands involved, more crises, and increased alertness in recognizing them and making them serve their social purpose.

The increasing frequency of crises is connected to the undervaluation of gold in the system; as we shall see, the simplest explanation takes us a long way. Suppose we start out with a system of currencies, goods, and gold and fix the price of gold in terms of currencies. Then allow a vast expansion of money, say, because of a war, which results in higher currency prices of all goods in the system except gold. Since gold is also a commodity, it will then be undervalued. Now take the most important of the currencies, the dollar, let all countries fix their exchange rates to the dollar, leaving the U.S. to keep the dollar convertible into gold at a given price. As long as the other central banks want to hold and accumulate both gold and dollars for use in their reserves, world inflation will have the effect of decreasing the residual supply available for central banks and thus drawing down (or preventing an increase in) the gold stocks of the U.S. more than would have been the case in the absence of world inflation. In a world where policy-makers act on slogans rather than analysis the problem will be misread as a balance-of-payments problem of the U.S. rather than as a problem of world inflation or the inadequacy of the reserve media.

To bring out the argument on a diagram let us restrict consideration to gold and three currencies: the dollar, the franc, and the mark. If the supply of currencies were in equilibrium at the existing price of gold and if exchange rates were at their equilibrium level, we could find an equilibrium price for the mark and the franc in terms of the dollar at which there would be no excess demand for any of the currencies or gold; this point would be illustrated by *Q* in the diagram where all three curves intersect. But if currencies had been supplied to excess and world commodity prices had risen, there would be no single exchange rate for the mark relative to the dollar or the franc relative to the dollar that would clear all markets simultaneously; the original equilibrium would be a position of excess demand for gold and excess supply of each of the other currencies. In such a case, exchange rates could be found that would equilibrate the French and German balances (point *A*); the French and American balance (point *B*) and the American and German balances (point *C*) but not all three balances simultaneously. The point *Q* would indicate deficits of all three countries and disappearance of monetary gold from central bank stocks. The point *R* would reflect the situation prevailing in November, 1968, neglecting, of course, the fact that the price of gold in the free market is separated from the official price. During the November, 1968, crisis the situation was at the point *R* and the authorities had a hard time deciding which, if any, pseudoequilibrium to move toward.

The main point I want to make with this diagram, however, is that crises are likely to result again and again when there is one price relationship that is not in equilibrium and all prices except that price can be changed. If gold is underpriced while the mark is undervalued and the franc overvalued relative to the dollar, exchange rate solutions can relieve the temporary pressure on one or another of the currencies until growth and an accumulation of demand for reserves create reserve scarcities for the next country. Faced with this situation relatively small disturbances can crack the system apart.

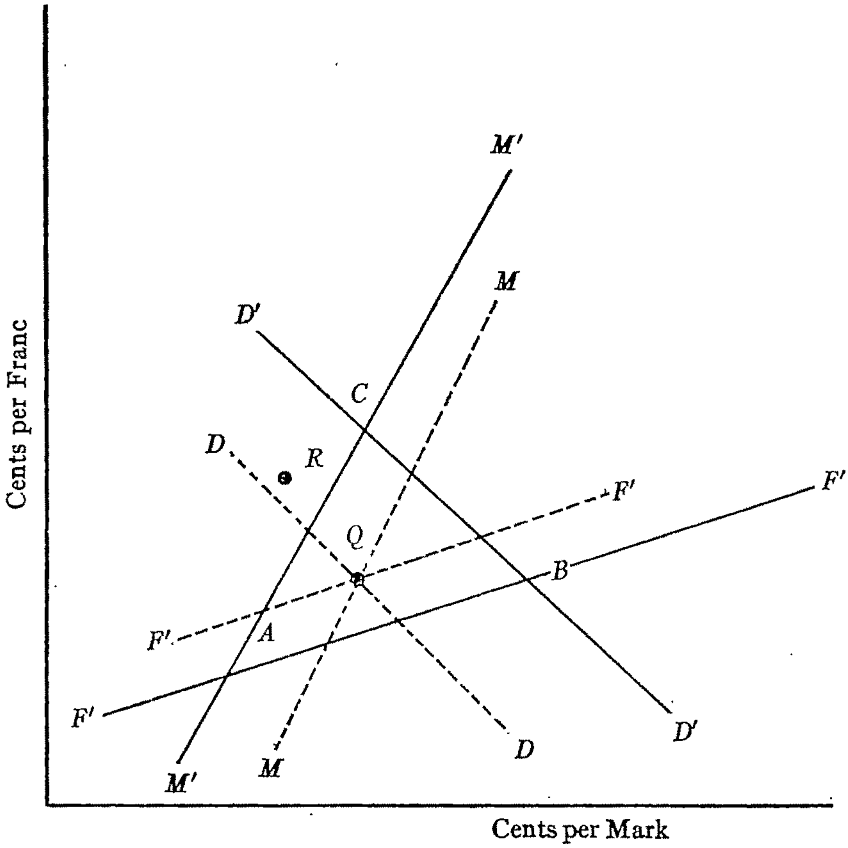


FIGURE 1

NOTE: The dashed lines, *DD, MM* and *FF,* intersecting at *Q* represent the initial equilibrium iso-balance-of-payments lines of the U.S., Germany, and France. Inflation with a fixed gold price has the effect of shifting these curves outward so that, at the original equilibrium *Q,* there would be an excess supply of all currencies and disappearance of gold into the private market. The system has no equilibrium until the real price of gold is raised either by price deflation or an increase in the currency price of gold; an alternative approach is to create a substitute, paper gold. The point *R* may be taken as the representative point in November, 1968.

There are two basic approaches out of the impasse. One is to raise the real price of gold; the other is to create paper gold. Now there are two ways of raising the real price of gold. The method chosen in 1930–34 was to deflate the world economy. This solution is rightly rejected universally by everyone. The other method is to raise the currency price of gold. Opinion on the merits of this solution is divided. During the November crisis a large number of economists favored a reduction in the mark price of gold. But there is a growing sentiment for a universal reduction in the par values of all currencies. I believe, however, that this solution is not necessary and is far inferior to the managed use of paper gold in the system. But before developing this argument it will be necessary to turn to the unique role played by the U.S. in the system and the malaise that is somewhat monotonously referred to as the U.S. balance-of-payments deficit.

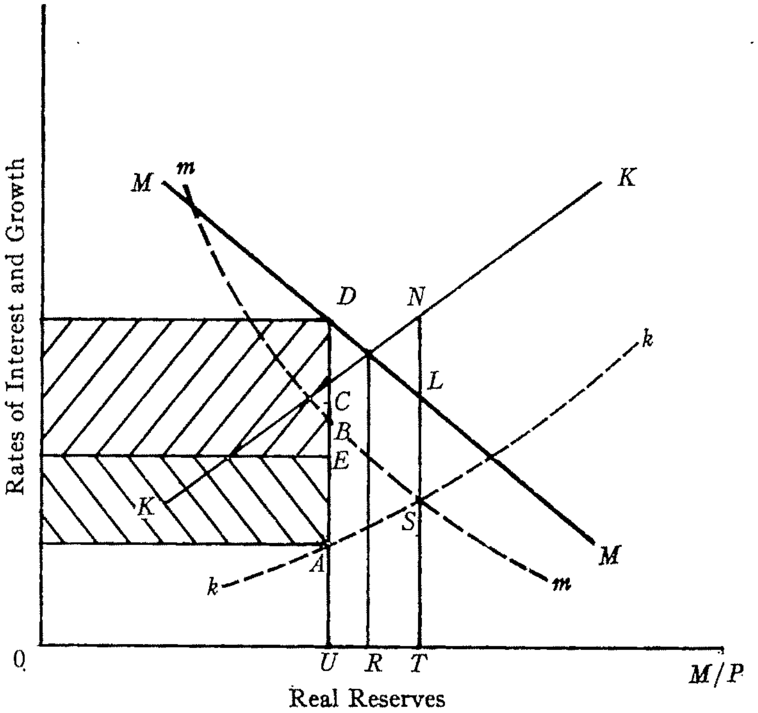


FIGURE 2

NOTE: Rates of change are plotted on the ordinate and real reserves on the abscissa. *MM* plots the money-balance schedule; *KK* the capital balance schedule. Rates of growth of money expansion are indicated by the vertical distance between *mm* and *MM;* rates of real growth by the vertical distance between *KK* and *kk*. In the nineteenth century the rate of real growth *NS* exceeded the rate of growth of gold production *LS* and deflation at the NL ensued. Since 1958, however, rates of money growth exceeded rates of real growth *CA* giving rise in rates of inflation *DC*. Gold production *DB* disappeared into private usage and rate of monetary growth *DA* was provided almost entirely by the U.S., *DE* representing U.S. gold losses, and *EA* the increase in U.S. dollar liabilities. The shaded areas give the absolute magnitude of the U.S. deficit.

II. *The U.S. Deficit*

In order to focus attention on the failure of the world economy to attain balance in the context of growth, let us analyze it as we would a closed integrated monetary system. Although this requires a leap of the imagination, it is fully justified in a world of fixed exchange rates and can indeed be derived from the aggregation of conditions of balance in a model in which an arbitrary number of national monetary systems are explicitly given full consideration. Monetary equilibrium in the system is dominated by the growth in the production of goods and services on the one hand and growth in the production of money on the other. Because it is a closed economy, there is no net lending outside the system so the excess demand for securities in it must be zero.

In Figure 2 real money balances are placed on the abscissa and various rates of change are placed on the ordinate. The curve *KK* shows the relation between real money balances and the real rate of interest at which the world community is willing to hold the stock of capital. The curve MM gives the relation between real money balances and the money rate of interest at which the world community of central banks is satisfied with the real stock of reserves balances it actually has.

If the world economy were not growing and no new money (e.g., gold) was being produced, equilibrium would settle at the point Q where MM and KK intersect. Real and money rates of interest would be equal (at the level QR) and the stock of real international money would be OR.

We must now take into account growth. At any moment of time the stock of nominal reserves balances is given, for the world as a whole, so that movements to the right and left on the ordinate implies lower and higher price levels, respectively. At each price level there will correspond a given rate of production of money and a given rate of production of goods. Draw kk to denote the given rate of goods production expressed as a proportion of the capital stock so that the vertical gap between kk and KK will indicate the percentage rate of goods production and therefore the commodity growth of the system. Similarly, draw mm so that the vertical gap between MM and mm indicates the rate of money production as a percentage of the money stock. Then equilibrium in the system will be established at the point where kk intersects mm; it is readily shown that S is the only position at which stocks and flows of goods and money are in equilibrium and that expected price changes are realized. At the equilibrium point S the desired and actual level of real money balances is OT; the real rate of interest is TN; the nominal rate of interest is LT; the rate of monetary expansion is SL; the rate of growth is NS; and the excess of the real rate of interest over the nominal rate of interest, NL, is the rate of deflation in the world economy. This configuration depicted at S might be taken to reflect the position of the world economy in the deflationary days of the gold standard between 1875 and 1895.

TABLE 1

Reserve Changes
1958–68*

	ROW	U. S. Deficit
Gold	+10.5	−9.9
F.E.	+ 8.3	−7.2
Total	+18.8	−17.1

* International Financial Statistics (Dec., 1958–June, 1968). Measure of U.S. deficit excludes swap arrangements, IMF position and foreign exchange holdings; the latter amounted to $2.5 billion in June, 1968.

But we are interested in current events, not past history. From the end of 1958 to the middle of 1968 world monetary gold stocks did not increase; virtually all new gold production went into private hands. But monetary reserves of countries outside the U.S. did increase because countries acquired dollars and gold from the U.S. which they earned by generating balance-of-payments surpluses. The rest of the world (countries outside the U.S.) bought both dollars and gold from the U.S. who became over the period the sole supplier of reserves leaving aside IMF Drawing Rights.[1] Enough reserves were bought to finance a gently rising price level throughout the world. Thus the situation characterized by the dollar standard as it operated from 1958 to 1968 can be best portrayed by a lower supply of liquidity (relative to capital, income, or trade), higher nominal interest rates, and perhaps real interest rates that were about the same as in the last part of the nineteenth century. In the diagram we can make the following average identifications for the period 1958–68:

---

[1] The position is not altered even if IMF drawing rights are counted as part of liquidity because the important part of the increase in Fund liquidity for the rest of the world was at the expense of the decrease in the Fund position of the U.S. Normally the gold component of an increase in IMF quotas is financed by transferring U.S. gold to the gold component of the IMF; although a bookkeeping device was invented to disguise this fact in the most recent increase in quotas in 1966.

$UD=i$, the nominal rate of interest
$UC=r$, the real rate of interest
$AC=\lambda$, the rate of growth
$AD=\rho$, the rate of monetary expansion in the rest of the worold
$DC=\pi$, the rate of inflation

The distance *DB* represents both gold production and gold consumption (disappearance into the private market).

ED is the gold sales of the U.S. as a fraction of world reserves.

EA is the annual increase in dollars held by foreigners as a fraction of world reserves. The shaded areas represent the U.S. deficit, divided between its two components: gold losses and the increase in dollars held by foreigners.

Of course Figure 2 is an oversimplified presentation of the world system. What two-dimensional representation could be otherwise? Its purpose is to dramatize certain relevant features of the world system and to emphasize how inadequately founded is that view of the world system which analyzes the U.S. on the same basis as any other country. The U.S. became the world supplier of reserves over the period 1958. There is no harm, of course, in speaking of a U.S. deficit any more than it would be harmful to say that the Federal Reserve System is in perpetual deficit because it pays out more dollars than it takes in. Words are harmless; it is the action taken on misinterpretation of the meaning of concepts that do the damage.

The crucial points to be emphasized can now be summarized. Because the world economy is growing countries collectively want to accumulate reserves. In order to accumulate reserves they try to run balance-of-payments surpluses. But they need an asset in which to hold their reserves. At present they can accumulate gold or dollars or some other national currency. But at the current price of gold there are no supplies of gold available to central banks. The U.S., however, has a special responsibility in the system. If other countries want gold they can take it out of the U.S. stock. Thus the U.S. is the only source of reserves in the system providing gold and dollars. This is called the U.S. deficit.

The only way the deficit can be corrected is by making available to other central banks an asset that is not provided by the U.S. The U.S. has been up to now the only country that is willing to supply the reserves other countries want. No other country than the U.S. was willing or able to supply gold on demand for its currency; and no other currency than dollars was needed for exchange transactions. This is the reason why the U.S. has run a payments deficit from 1958–68, and why, in the absence of paper gold, it will continue to run deficits.

III. *Paper Gold*

I now turn to paper gold. The demand for additional dollar holdings will continue but only in connection with another asset in the system. Real gold is no longer adequate to assure sufficient and sustained reserve growth. A paper gold solution is the only approach that can create the conditions for rational world monetary management, end the systematic crises, and end the U.S. balance-of-payments deficit.

There are several ways of creating paper gold. One is to create a right to draw on the IMF and give liabilities incurred a gold guarantee, as was done with the Special Drawing Rights. This does not meet the basic world monetary problem; the instrument does nothing to solve the confidence problem; and the supply of

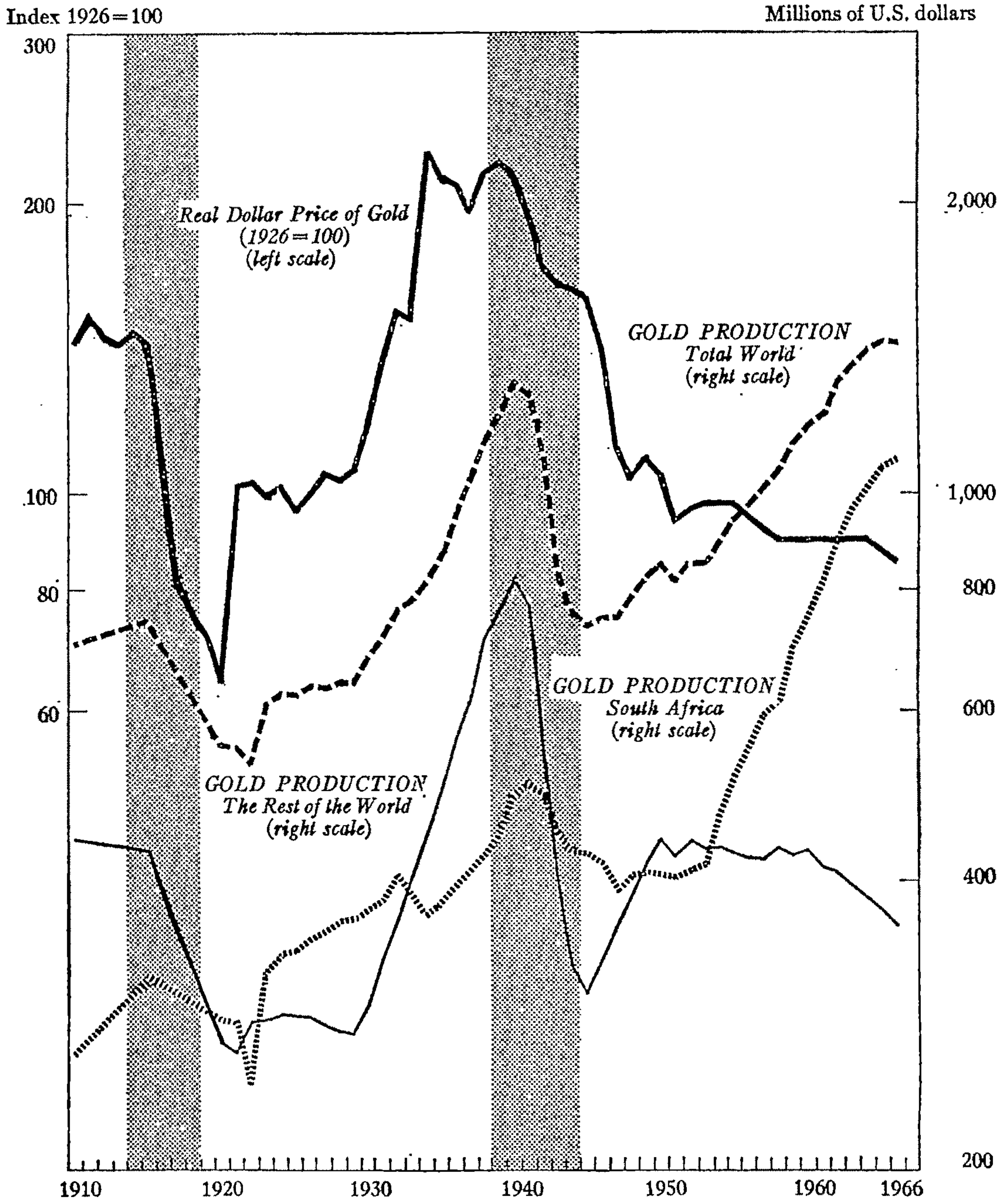


FIGURE 3

GOLD PRODUCTION AND A REAL DOLLAR PRICE, 1910–66

Source: Fred Hirsch, "Influences on Gold Production," *IMF Staff Papers* (Nov., 1968).

SDR's that will be created will be inadequate. Over the next ten or fifteen years the world will need perhaps $50 to $80 billions of new reserves but my prediction is that the nations will not be willing to create even half of that amount of SDR's.

The best solution, in my opinion, is to create paper gold by centralizing reserve assets. A number of plans for centralizing reserves have been discussed; most have parents in one of Keynes's or Triffin's plans. My own preference is to put all reserves in an international monetary pool (IMP) and use the certificates members get in exchange (intors) as world money. Since the March communique they could use a shadow price of gold as a new instrument. Suppose, for example, the countries contributed to the pool the following reserves: gold, 1 billion ounces; foreign exchange, $25 billion; other, $5 billion. The IMP would then have to decide how to value gold. Since the official market has been separated from the private market one could make a virtue of this separation and use the official price of gold as a regulator. Let us suppose that initially gold is valued at $35 an ounce. The initial balance sheet of the pool would be as follows:

The IMP

Assets		Liabilities	
Gold	$35 b.	Intors	$65 b.
Other	$30 b.		
	$65 b.		$65 b.

Now suppose the members decided to increase the supply of intors by $3 billion a year for the next five years. They could acquire more gold, foreign exchange or SDR's, creating intors in exchange. But this is no longer necessary. They have the other alternative of changing the value they place on the million ounces of gold. If they did not acquire any additional dollars or new physical gold stocks, the additional intors could be created through an increase of $3.00 an ounce in the gold price resulting in, at the end of the next year, the following balance sheet:

The IMP

Gold	$38 b.	Intors	$68 b.
Other	$30 b.		
	$68 b.		$68 b.

Eventually, gold would be worth more as money than as a commodity and thus establish the two-price system on a basis consistent with Gresham's Law. The free market price of gold could be dominated by the gold policy of the IMP. The major intervention currency, the U.S. dollar, would cease dealing in gold and deal only in intors to which the dollar would be fixed in value, and the exchange rate system would continue as before.

The essential point to notice, however, is that it is not tonnage that is required to make new international money. Paper gold is better than real gold.

Conclusion

I have now argued (1) that increasing crises occur in the system because of the undervaluation of gold, (2) that the elasticity of the gold exchange standard based on the U.S. balance of payments has enabled the system to operate successfully in the past decade, and (3) that the use of paper gold is the only device on which a lasting world monetary system can be based.

SPECULATIONS ON GOLD SPECULATION

By Fritz Machlup
Princeton University

"Speculation on speculation is so speculative, there is no point in it." This might be a remark by one of my critics, but in fact it is a quotation from a press statement by Prime Minister Keith Holyoake, of New Zealand. He was not commenting on the title of my paper but replying to questions about possible terms for the bombing halt in Vietnam. His three uses of the word speculation related to mental operations, none to market operations. The object of my discussion, however speculative, is gold speculation.

Since I am known to have said nasty things about the gold lobby, I want to reassure you that I am not going to be rough on speculators. Speculation in all sorts of markets is usually a "good thing," especially when the speculators make money. As a rule, their profits can be taken as proof that they were right and that their actions were in the national interest, even if the results were irritating to government officials who found some of their policies revealed as ineffective, inexpedient, or injurious. People who expect certain developments in the production or consumption of certain commodities, or in the earnings of certain corporations, or in the exchange value of certain currencies, often act upon their anticipations by speculative purchases or sales of the commodities, securities, or currencies in question. My studies in welfare economics have not led me to castigate the speculators. We all can understand, of course, that devotees of the green-eyed monster may be filled with envy of the successful speculators and with *Schadenfreude* about those who lose their shirts in the process. Since gold speculators have fared badly and may end up even worse, I wish to state that I am sorry for them and hope they will choose better financial advisers in the future.

Perhaps I should show how badly those have fared who have bought and held gold for several years. Those who bought in 1957 had, at the end of 1967, ten years later, only about 31 percent of the investment that a man had who at the same time bought in the New York Stock Exchange a mixture of the 425 industrial common stocks used in the Standard & Poor index; those who bought in 1960 had, in 1967, seven years later, only 48 percent of the net worth that the stock purchaser had; finally, the gold buyers of 1962 had, in 1967, five years later, only 53 percent of the net worth of the same original investment in stocks. To be sure, the free-market price of gold has risen since the end of 1967 by between 6 and 12 percent, but stock prices have risen even more, so that the advantage to the holder of stocks is even more conspicuous. The gold holders have also done worse compared with dollar bond holders, compared with holders of dollar savings deposits, yes, even compared with holders of British securities or with holders of sterling deposits despite the devaluation of the pound. There are few assets that have done worse than gold as a store of value.

Some diehards say that he who laughs last will laugh best and that one nice weekend we shall learn that the price of gold has suddenly been doubled, evidently as a result of black magic or blackmail applied to a reluctant government of the United States. This has in fact been the pious expectation of the speculative buyers of gold for the last fifteen years. If this pipe dream did come true, those who bought their gold more than five years ago would still be behind those who bought an average mixture of American industrial stocks, though they would definitely be ahead of the holders of savings deposits, especially in pounds sterling. To evaluate the chances of the speculators' dream coming true is one of the tasks of my discussion.

Who Is a Speculator?

Before we can go on, we have to settle the question of when a buyer of gold may be called "a speculator." I would never have thought that this could be in doubt, but I found I had been wrong. The Annual Report of the Bank for International Settlements, for example, denied that purchases of gold were speculative when they were made out of the buyers' savings or when the gold acquired was not resold within a brief period—even if the total of purchases was between 50 and 100 percent above that in the preceding year.

The word "speculation" may mean different things in the trade jargon of different markets. In the securities market one buys for "income" or for "appreciation," or one buys stocks with narrow or with wide fluctuations in price, and the designation "speculative" may be given to the second alternative in each pair. In the foreign-exchange market one may be in a covered or in an uncovered position, and respectively be regarded as hedging or as speculating. On the other hand, one who speculates on a depreciation of his own currency and holds foreign balances (in an uncovered position) may explain his actions as hedging against inflation. This is hedging in a different sense, for in this case speculating becomes hedging, in the parlance of the operators. In the commodities market purchases not for use within the "normal" duration of holding inventory are regarded as both speculative and "hedges against a rise in price"; on the other hand, the term speculator may be reserved for a purchaser of futures who never takes delivery but sells his rights before the due dates.

It is possible that the gold market has its own trade language, in which the word speculation is given a restrictive meaning; for example, purchases with borrowed funds or purchases for resale within six months or a year. For economic analysis this would be of little use. The essence for purposes of market analysis is, first, that the gold purchased is not processed (made into watchcases, bracelets, wedding bands, tooth inlays, bridges and crowns, parts in electronic equipment, linings in space missiles, and so forth), second, that the purchase is made in expectation of an (absolute or relative) increase in the price of gold, and, third, that the gold is for potential or eventual resale when the price is "right" or when expectations change. This may be in a year or two, or it may be in ten or twenty years. For the definition we select the second of these three conditions and thus decide that any purchase of gold in expectation of an increase in its price shall be characterized as speculative.

Tradition, Safety, Fancy

There is the case of the traditional hoarder of gold, who buys gold irrespective of actual or expected price and holds on to his gold no matter what happens to the price. Perhaps this nonspeculative hoarder is only an ideal type, who does not exist in reality, but we want him anyway as a limiting case. For there may be a whole spectrum of gold holders, ranging from one who will sell when the price rises by 5 percent to the one who will never sell, even if the price rises a quadrillion times. Some experts assure us that such incorruptible hoarders really exist: Indian maharajahs are given as illustration. One may wonder whether all but these never-never-to-sell gold holders should be called "speculators" or perhaps only those whose reservation prices may become practically relevant. The question, however, is ill-conceived, for it focuses on the decision to sell rather than on the decision to buy. In order to decide whether a purchase is or is not speculative we have only to ask whether it would have been made also in the absence of any expectation of an increase in the price of gold. The Indian maharajah (we may say in order to establish the type) would have bought, but the Arab sheik not. Hence, the former is a traditional hoarder, the latter a speculator.

My definition of speculation would cover purchases of gold by people who look for a "safe investment" of their savings. For, obviously, they regard gold as a safe investment in the sense that they trust it more than assets denominated in currency; this means that they think there is a good chance for money to depreciate in terms of gold; that is, for gold to appreciate in terms of money. Hence, they buy gold for appreciation and, since they cannot benefit from such appreciation unless they sell eventually (perhaps when no further gain is expected), their gold is for potential resale. Gold purchases as safe investment are therefore speculative in that they presuppose expectations of an increase in price.

A very different motive for buying and holding gold is for "safekeeping" against confiscation, taxation, and theft. In financially undeveloped countries many people, not only the rich, prefer to hold gold because of the ease of hiding it. Expectations of appreciation may be relatively, or completely, irrelevant; what matters to the owners is protection from total loss. And what matters to the analyst of the gold market is that the gold held for safekeeping (in vaults, under mattresses, in holes in the ground) is not likely to reappear on the market, at least not in response to price changes. This gold is dishoarded when the owner needs money for urgent purchases or payments, or when the owner dies and his heirs want to liquidate their inheritance. Since predictions regarding the reappearance and transfer of such gold are altogether impossible, it may be expedient to regard these stocks as hoarded "for keeps."

Some worshippers of gold as the noblest and most beautiful metal have delivered veritable elegies on the pleasures from treasures of gold in coins, bars, and other shapes. That the happy possessors enjoy the sight of their treasures is not inconsistent with speculation for gain from eventual sale. This twofold motive for buying and holding gold prevails often in the acquisition of art objects. A connoisseur may have bought, before 1910, several paintings by Matisse for both aesthetic and pecuniary appreciation; he was pleased while they were hanging on his walls, and pleased also when he realized a large capital gain upon selling them fifty years later. As he considered the pictures a good investment, he was a speculator, even if for many years he derived a psychic income in terms of visual enjoyment and social prestige. Had he kept the paintings in storage until he sold them, it would of course be operationally more evident that he had bought them for speculation.

Our purpose in these semantic or conceptual observations is analytical. For the interpretation of statistical data showing private purchases of gold over a series of years it matters a great deal whether some or many of the purchases were speculative. A discernible trend can reasonably be projected into the future if speculation has played only a negligible role; such a projection will be misleading, however, if the role of speculation has been substantial. Since speculators buy in order to resell in the future, market analysis calls for inquiries to ascertain the share of speculation in past acquisitions.

Fundamentals of Market Analysis

In the case of a perishable commodity traded in a competitive market, supply is derived from cost of production and demand from utility of consumption. For nonperishable, storable commodities, supply is composed of new production and reductions of stock, and demand is composed of current consumption and additions to stock. For commodities controlled by firms or groups of firms with considerable monopoly (monopsony) power, sales (purchases) will to a large extent be determined by sellers' (buyers') anticipations of the effects of their sales (purchases) upon price. It is probably clear that gold is an eminently storable commodity, with stocks of a size many times annual output and still more times annual consumption; that a very large part of the total output is under monopoly control; and that a large part of the potential demand and a large portion of the total stocks held are controlled by a group of cooperating authorities. In these circumstances analysis of the gold market is devilishly speculative.

To proceed systematically, we shall begin with a schematic tabulation of the components of supply and demand in the gold market.

Estimates of production and sales, for a series of years past, are sometimes confused with market analysis. Of course, such estimates are the prerequisites of analysis, but naïve projection of past series of data by assuming the observed rates of growth to continue into future years surely does not merit the name of analysis. Needed are reasonable assumptions linking each of the components to factors about which information is or may become available. Where no information can be obtained, the lack of it has to be noted and the possible influences and developments have to be considered in general terms—and thus the analysis cannot help being speculative. Each of the components of supply and demand may be subject to change: production may change because of technological factors, consumption because of income and substitution effects, private stocks because of changing expectations, and monetary stocks because of changing policies and institutions.

THE GOLD MARKET

Supply	*Demand*
Current Production	Current Consumption
From old mines	In jewelry
newer mines	dentistry
newly discovered mines	industry
improved techniques	
ocean shelves and ocean water	
artificial gold	
Scrap	
Reductions of processors' inventories	Additions to processors' inventories
Reductions of other private stocks	Additions to other private stocks
As funds are needed	For beauty, prestige, tradition
price is high enough (profit taking)	safekeeping (from confiscation, theft)
price will not rise (getting impatient)	store of value (better than other assets)
price starts falling (getting bearish)	appreciation over years
	quick appreciation
Reductions of monetary stocks	Additions to monetary stocks

The possibilities and probabilities of factors capable of generating change will now be considered.

Production

In each of the four years 1964 to 1967, gold production in member countries of the International Monetary Fund exceeded $1,400 million at $35.00 an ounce. In 1964, the Soviet Union sold $450 million and in 1965, $550 million worth of gold. Assuming that these sales roughly equalled Soviet output, world production may be put at between $1,900 and $2,000 million per year.

South Africa is the largest producer, with over $1,000 million worth of gold per year. Most of this is from newer mines, discovered in the early 1930's and operated at relatively low cost—about $18.50 an ounce in 1967. In old mines, production cost is much higher, nearly $32.00 and some marginal mines may be shut down eventually. Closed mines exist in several countries, including the United States—gold mines once profitable, now dead and buried. Their demise is lamented by sentimentalists and by interested parties, who advocate their revival by means of government subsidies.

Forecasts about future outputs of the newer mines are contradictory. Some predict a gradual decline and point to the statistic showing that South African production actually declined from 1966 to 1967; but they usually fail to point out that this decline was by only 1 percent; that is, less than three days' output. Another decline may be recorded in 1968, for it has been reported that a mine was flooded by subterranean waters and operations had to be temporarily stopped until the water was pumped out. Other observers, not associciated with the Chamber of Mines, believe that production in the newer mines, which had increased steadily from 1950 to 1966, has not yet reached its peak.

Several possible developments may result in larger outputs of gold: (1) New gold deposits may be discovered. (2) Improved techniques of mining and refining, already introduced at several places, may be extended to many more mines. Their adoption requires investments, which mining companies now prefer to postpone but may in due course find attractive. (3) Techniques of extracting auriferous ores from the bottom of the sea, particularly from ocean shelves close to land, may be developed. (4) Techniques of recovering gold from ocean water may be developed. (5) Techniques of chemical transformation of other elements into gold (alchemy), now prohibitively expensive, may eventually become sufficiently cheap. It would be foolish to expect all five of these developments, but two or three of them may well occur within the next ten or twenty years.

In the meantime, however, labor cost is rising steadily and one may ask whether this would not lead to substantial cutbacks in output. In the older mines, this may well happen, but it is less likely in the new mines, where average cost could rise by 80 percent and still be sufficiently below $35.00 an ounce. Even there, one should expect some reductions in the volume of output, as an increasing marginal-cost curve is shifted upward with the rise in labor cost.

Consumption

In superficial analyses total private purchases of gold are sometimes designated "private absorption" as if it made no difference whether the gold purchased is made into jewelry, filled into dental cavities, held in hoards by Indian princes, or stored in Swiss vaults for potential resale. Such differences, however, are essential. We shall first discuss private "consumption" of gold; that is, changes in its shape that would exclude it from being traded in wholesale gold markets and may involve substantial complementary inputs of labor and other costs. For example, gold made into a necklace is not traded in the gold markets of London or Zurich, and the value of the necklace may substantially exceed that of its gold content.

Exact information on annual amounts of private consumption of gold is not available. Only twelve countries publish statistical data on purchases of gold by processing crafts and industries; the accuracy of these data has been questioned; and even if they were accurate, the published figures refer not to actual use but to purchases inclusive of quantities added to inventory. There can be no doubt of such additions to inventory, not only on the basis of "a priori reasoning"—since "typical" businessmen stock up when they expect the price of a material to rise—but on the basis of the published figures, which show reduced purchases in years following unusually large increases. (Such reductions were reported in the United States for 1961 and 1963.)

Purchases of gold, for use and inventory, by "industry and arts" in the twelve reporting countries were recorded at a figure of $502 million for 1966. Ten years earlier the figure was $165 million. The large increase has been explained by: (1) technological developments in industry requiring the use of gold in electronic apparatuses, space missiles, and other pieces of equipment; (2) the elasticity of substitution of gold for other metals, the prices of which have been rising; and (3) the income elasticity of demand for jewelry and other things made with gold. The first factor has no doubt been operative, but the increased use of gold in industry cannot account for much of the increased purchases by licensed dealers and processors, because it started at such a low figure that even a phenomenal rate of growth would not do the trick. The second factor is probably of minor importance, because alternative metals are either much more expensive or very much cheaper, so that the substitution effects of price changes are likely to be small. The third factor is certainly effective and important, since rising incomes, aggregate and per head, raise the demand for all (except "inferior") goods. We must, however, not lose our sense of proportion and naïvely accept the hypothesis that a 74 percent increase in total money income or a 48 percent increase in total real income in the United States from 1956 to 1966 could account for a 200 percent increase in purchases of gold. I conclude that an increasing share of purchases for inventory accumulation must have been a very substantial fourth factor in the increase of total purchases.

Jewelry, dentistry, and industry are the users of gold. Estimates for recent years have jewelry take three-fourths of total gold consumption, with dentistry and industry taking one-eighth each. This is why the first factor in the explanation of the increase in consumption—more uses in industry—cannot have been of great importance in the past. It may, of course, have increasing importance in the future. In order to predict future consumption we ought to have an idea of the income elasticities of demand. I have tried to get such an idea from the statistical series of the United States but have not succeeded. Gold purchases by "industry and arts" declined in several years from the level of the preceding year; in the nine years from 1951 to 1959 they were lower than in 1950; and an attempt to smooth the series by computing moving three-year averages was of no help, since it still exhibited several reductions from one year to the next. The effects of fluctuations in inventories have probably been too strong to be eliminated as long as actual data for inventories are not available. When purchases jumped from $102.2 million in 1963 to

$168.0 million in 1964, the notion of explaining such a jolt by the income elasticity of demand must obviously be rejected.

The lack of statistical or econometric estimates of the income elasticity of demand for gold in artistic, dental, and industrial uses is a serious obstacle in judging future increases in gold consumption. If users' purchases during the last few years included much stockpiling, it would be quite possible for such purchases in 1969 and 1970 to be below those of 1966 and 1967. After the inventories are worked down to nonspeculative levels, purchases of gold for processing will of course resume an upward trend, but whether the annual rate of increase will be 4 percent or 8 percent a year, nobody can tell. I doubt that the income elasticity of demand for the whole world exceeds 2. If world consumption by 1970 is $750 million (at $35.00 an ounce), which was the estimate of processors' purchases in 1966, and if total real incomes rise by 4 percent a year, we should reckon with an increase in gold consumption by 8 percent, or $60 million, to 1971. This order of magnitude of the increase is quite different from the "trend" which some statisticians believe to have spotted from the record of the past years. At this rate of growth it would take us into the 1980's until consumption can catch up with the present rate of annual production.

Of course, purchases may exceed consumption if people want to accumulate stocks. We shall deal with this question presently, but ought to comment first on another item: the availability and use of scrap.

Scrap

The statistics of sales, purchases, and use of gold are usually silent on the question of scrap. This is strange, since the use of scrap is not a negligible part of the total use of gold. According to at least one source, scrap provides almost one-third of the gold used by processors in the United States. Scrap may either be reused in the places where it is collected or may be offered for sale to other users. It is not clear whether the purchase of scrap is or is not included in the published statistics, but in all probability it is not. This may imply underestimates of both supply and demand.

All three components of gold consumption yield scrap and all three use scrap. In jewelry, old-fashioned pieces, trinkets, and ornaments provide the bulk of the scrap, to be refashioned into pieces of jewelry suiting present-day tastes. In dentistry sordid Nazis in the concentration camps organized recovery of gold from the teeth of corpses; less exceptionally, dentists sometimes save the dust derived from smoothing and polishing inlays, crowns, and bridges prepared for the teeth of their patients. In industry, shavings from gold parts fitted into various apparatuses and gold dust accumulating in the operation of certain kinds of equipment are collected for resale.

Ignorant of the magnitudes in question, I can only suggest that here seems to be an item to be looked into, though it may prove to be negligible after all.

Processors' Inventories

The importance of processors' inventories has been shown above in the discussion of the statistics of purchases of gold. In the tabulation of components of supply and demand, additions to inventories are shown on the demand side, reductions of inventories on the supply side. It should be clear, however, that only the former give rise to market transactions; that is, to purchases of gold by processors. Reductions of inventory, on the other hand, do not result in sales. They call attention only to the necessary correction of the figures for consumption of gold, since some of the gold used need not be purchased but is taken from inventory. The point of all this is that processors' purchases are larger than their actual use of gold in years when price is expected to rise, and smaller than actual use in years when such expectations are no longer prevalent.

In an article published a few months ago,[1] I estimated the size of processors' inventories to be between \$500 and \$1,000 million; that is, between one- and two-years' consumption of gold. This estimate applied to March, 1968—the end of the speculative accumulation of inventories.

Other Private Stocks

My estimate of the stocks accumulated by other private holders was \$20,000 million, of which I thought \$15,000 million would be for potential or eventual resale, depending chiefly on expectations regarding the price of gold.

The distinction between gold acquired with the intention never to resell and gold acquired for eventual resale at a profit is important, partly because it aids in the analysis of demand, but partly also because the motives in buying may remain valid in influencing the decisions to sell, which is relevant for the analysis of potential supply.

In my synoptic table, in the list of motives for private stockpiling by nonprocessors of gold, the first entry refers to the hoarder who has no intention of reselling ever. He wants to possess gold in order to enjoy its glitter or the prestige which its possession may secure, or perhaps just in order to continue an old tradition. He is impervious to any changes in the market value of his gold and indifferent to any prospects of the price rising or falling in the near or distant future. Of course, the intention at the time of purchase of holding this gold forever does not bar the possibility of some being sold out of such hoards. The owner may find himself in urgent need of liquid funds and thus compelled to sell; or the owner's heirs may not have inherited his determination never to dishoard.

The next entry in the list of motives is safekeeping of one's wealth against confiscation or theft. Few marketable material things have as high a value per cubic volume as gold; hence, a very small hole in the wall or in the ground can hold a very large amount of gold, valued in money or other readily exchangeable things. Thus, in countries where people are fearful of thieves, official or private, many owners of wealth, even quite modest wealth, prefer to hide some gold in places where it cannot easily be found or got at. Those who acquire gold for the purpose of keeping their wealth safe against theft or confiscation do not necessarily intend to hold it forever. They may intend to dishoard it in later years. But these hoarders of gold, so we are told, are not significantly influenced by expectations regarding its price. They would buy gold even if they did not expect it to appreciate in terms of domestic money; indeed, they would buy gold even if they knew it might lose some of its value, if only they could feel safe from losing all to thieves and robbers.

The demand for gold by both traditional and hide-away hoarders is probably a function of income. Hence, the increase of this demand may be largely determined by its income elasticity. Unfortunately, we know little about the magnitudes involved: we do not know how much gold has been acquired annually by price-insensitive hoarders, how much their incomes have been, at what rates these incomes are increasing, and how much of these increases is spent on acquiring more gold. Guesses thrown around by the allegedly informed about the gold purchases by this type of hoarder vary between $100 million and $300 million for 1966 or 1967. (I have no idea how much of the stories about price-insensitive hoarders I should believe.)

---

[1] Fritz Machlup, "The Price of Gold," *The Banker*, Sept., 1968, pp. 782–91.

One of the various factors that might change the propensity to hoard gold would be an increase of attractive investment opportunities in the countries in question. If organized capital markets are developed, if securities can more easily be bought and sold, if confidence in securities is strengthened, if the political climate and legal institutions become more stable, the asset preferences of many people in many parts of the world may change. They may become holders of securities instead of hoarders of gold. This change can, of course, not come overnight and we may have to anticipate several years in which gold will be purchased for additions to private hoards.

We now turn to the speculative motive for buying and holding gold. In recent years the bulk of gold purchases for private stockpiling has been motivated by expectations that gold will appreciate in terms of money. This holds, we may recall, even for those who have bought gold as a safe investment or as a store of value, since such confidence implies that gold is expected to maintain its value more securely than money or securities; that is, an expectation that the money value of gold will increase.

Most of these buyers of gold, seeking merely protection against inflation, do not regard themselves as speculators; but for purposes of analysis it makes no difference whether a purchaser wants to avoid loss from the depreciation of money or whether he seeks capital gain from the appreciation of gold. As he is more conscious of his goal of seeking safety from loss through inflation, he usually has no specific expectations as to the timing and extent of the increase in the money value of gold. While he believes that gold is safer than currencies or securities, he does not speculate about whether its appreciation will begin next month or next year and whether it will occur in one fell swoop or in several smaller jumps.

The last two motives on my list are those of conscious seekers of gain from appreciation, either in the long run or in the near future. These buyers are influenced chiefly by rumors, news reports, financial analysts' releases, and bankers' advice. They buy gold because they have been told that its price will be doubled or tripled, either because the monetary authorities of the United States will find it expedient or imperative to take this action, or because gold production is going to be reduced and/or gold consumption is increasing at a fabulous rate each year, or because all other commodities have had their prices more than doubled since 1934 and the price of gold will necessarily catch up with the rest. Whatever the reason, they expect an increase in the price of gold and they do not want to miss the chance of a good profit.

The expectation entertained at the time of the purchase may or may not remain unchanged. In some instances—for example, when a doubling of the price was expected to occur within a very short time—the expectation is disappointed and has to be revised. If a higher price—say, an increase by 50 percent—were expected, but no specific time of its occurrence, the expectation may be maintained for years. If the market price of gold actually begins to rise, the holder's original belief may be strengthened and he may be determined to hold out until the price reaches the expected level; or, as that level is approached, he may become even more optimistic and raise his anticipations and his reservation price; alternatively, he may become fearful that he had been overconfident and may prefer to take his profit, preferring a bird in the hand.

Only in the case of invariant expectations can one conceive of a supply curve depicting a set of reservation prices and quantities offered from the speculators' accumulated stocks. Yet, this would be a rather unrealistic conception of the supply of gold from reductions of private stocks. Changes of expectations will as a rule dominate the situation. I have listed three typical circumstances as the ones most likely to affect gold holders' decisions to sell. If the gold market registers some price advances, holders may decide to sell because they think that the price has increased to a level from which a decline is more probable than a further rise; they take their profit, however modest. If the price of gold fails to rise substantially and keeps hovering closely above the present level, many holders may get tired of waiting for the great increases which they originally expected; they may give up and sell for the current market price. Finally, if the market shows a series of price reductions and it is decided that the price will not be supported by the monetary authorities, many holders may get scared of a serious crash and sell as fast as they can, thereby causing the crash they feared.

Most of the expectations of private buyers, holders, and sellers of gold are focused on what the monetary authorities are likely to do. Since the authorities have over the years been the largest purchasers of gold, are now the largest holders of gold, and can in the future be the largest sellers of gold, their role both in the expected and in the actual developments of the gold market undoubtedly dominates the scene. Their utterances, however casual and unpremeditated, may make speculators jump and cause waves of buying or selling in the private gold markets. Early activation of the plan to create and distribute Special Drawing Rights would probably indicate to speculators that the international monetary system can do without further accumulations of gold reserves, and this might seriously weaken the speculative motive for private buying and holding of gold.

Monetary Stocks

The gold stocks held by the monetary authorities of the member countries of the IMF and by the international authorities were $40,500 million as of June, 1968. Between 1949 and 1965 the authorities had purchased $8,800 million, or an average of $550 million a year. From December, 1965, to March, 1968, the authorities, meeting the demand of private stockpilers, had sold $3,000 million from monetary gold stocks, most of it in the last three months of that period.

The official purchases before 1965 and sales afterwards might be regarded as characteristic for a period that came to an end on March 17, 1968, when the active members of the gold pool agreed to stop selling or buying gold in the market and appealed to all other members of the IMF to do likewise. For several reasons, however, it would be wrong to assume that the insulation of monetary gold reserves from the free gold market would be fully effective and permanent. It would require that none of the existing monetary gold is sold in the free market and none of private holdings or newly produced gold is acquired by the authorities. The danger that some small central banks may break ranks is not serious. Serious, however, is that the Reserve Bank of South Africa has been buying the entire output of that country's mines, so that the statistic of monetary gold stocks shows and will continue to show a steady increase. Moreover, the Reserve Bank of South Africa has sold some gold in the Zurich market, a relatively small amount and only in small doses because the Bank wanted to avoid a serious price decline. Certain authorities, including the United States, would prefer that the South Africans sell all of the current output in the free market. This could drive the price below $35.00 an ounce, since consumption of gold is still far less than production and since speculative purchases have stopped.

The South Africans, who for years have tried to persuade the financial world and the monetary authorities to double the price of gold, are now appealing for the resumption of official purchases of their new gold production at the long-established support price of $35.00. They point to a legal right, based on the Articles of Agreement of the Fund. For, when South Africa has a deficit in its balance of payments, it may, like any member, make a "drawing" and purchase from the Fund dollars or other convertible currencies, paying with rands, its own currency. Then it may repurchase its currency from the Fund with gold. This would, of course, amount to a sale of gold to the Fund. South Africa might thus enjoy a two-price system, selling small portions of its production in Zurich for a better price and the rest to the Fund at $35.00. (A firm in the United States, using its monopoly power in a similar fashion to practice price discrimination, would be prosecuted for violating the antitrust laws. This may explain the moral indignation of some American officials about the insistence on a support price of $35.00 for gold withheld from a higher-price market.)

Perhaps I should not engage in idle speculation about the way this conflict will be resolved. It seems clear, however, that many central bankers want to maintain the gold myth, the widespread belief in a permanent intrinsic value of gold. If the entire production of gold had to be sold in the free market, the price would surely drop far below the official exchange value of $35.00, and this could damage the popular faith in gold. Moreover, governments in countries where private holdings of gold are large want to protect their people from an unnecessary loss of wealth. Thus, one may expect that, at least for some time, a number of governments will use their influence in favor of a scheme that provides a floor for the market price of gold. Such a scheme would require further additions to monetary gold reserves in amounts insufficient to secure adequate increases in aggregate monetary reserves, but sufficient to save the excess supply of gold from going begging and depressing the price. The situation may change after several years, when consumption will possibly catch up with production and the required support from monetary authorities be changed from buying to non-selling: they would not have to buy an excess supply but would merely have to refrain from selling off their reserves too quickly.

Incidentally, I cannot help finding it more than strange that the same men, in or out of government, who most strongly believe in or pronounce on the eternal and intrinsic value of gold also most strongly oppose the removal of the artificial floor under the market price of gold. They call for continued support of the gold price through purchases by national or international monetary authorities. If the value of gold were really immanent and permanent, there would be no need for monetary institutions insuring an elastic demand for any excess supply of gold in the market. The truth of the matter is that under present conditions the price of gold in the free market cannot be maintained at anything near the present level if the authorities refuse to purchase gold and to hold on to the gold they possess. Perhaps the authorities will continue to keep all their gold and even to add to it, but they ought not to kid us and should recognize that gold is a safe asset only because central bankers act as if it were a safe asset.

The Present Situation

I shall conclude with a brief sketch of the present situation in the free markets for gold. Consumption running at an annual rate of perhaps $500 million and nonspeculative hoarding at a rate of $200 million at most, a total of $700 million, calculated at $35.00 an ounce, is being absorbed. As the Russians continue to withhold their output, sales from new gold production are confined to the relatively small amounts sold by South Africa, say $300 million, and another $300 million produced by other countries. This is less than current consumption and hoarding; the gap is probably being filled by reductions of private stocks accumulated during the days of the gold rush.

If the South Africans increased the amounts of gold released to the free markets, the price would fall; if they reduced them, the price would rise. Were they to sell their entire output in the market, the price would tumble; were they to stop all sales, the price could soar to higher levels. A sharp drop would almost certainly be aggravated by sales from speculative holdings; a good rise might be reinforced by a resumption of speculative purchases. This is a highly unstable situation and the South Africans may take credit for their efforts in stabilizing the market by doling out their gold with discriminating monopolistic restraint. They would be glad, I suppose, to leave the responsibility for stabilizing the price of gold to the monetary authorities. And that this is where the responsibility lies and that a stable price of gold is imperative for the prosperity of mankind, these are articles of faith among the traditionalists in the world of finance.

THE GOLD RUSH OF 1968 IN RETROSPECT AND PROSPECT

By HARRY G. JOHNSON
London School of Economics and *University of Chicago*

The gold crisis of March, 1968, which led to the establishment of the present "two-tier" price system for gold, was touched off by short-run speculation on a rise in the price of gold, derived from uncertainties generated by the devaluation of sterling, the extraordinary deterioration of the U.S. balance of payments in latter 1967, and growing disbelief—justified in the event—in the willingness and/or ability of the gold pool countries to hold the market price at $35 an ounce. Given the unpopularity of successful speculators on financial disaster among both central bankers and academic economists, there is a natural temptation in reviewing the present situation to concentrate on the limitations on the power of the speculators to absorb gold, and the capacity of the world's central banks to defeat them by selling gold. According to this view—which I confess I myself expressed at the time—the March decision to break the link between the monetary and the free-market price of gold represented a failure of nerve on the part of the members of the gold pool, and the present problem is to show the speculators the error of their ways by manipulating the free market price down to or below the monetary price, while proceeding towards the ultimate demonetization of gold by introducing the Special Drawing Rights and supplementing them by additional schemes for funding existing reserve assets into international credit reserve money. This is what the logic of the situation suggests, as is I think agreed in general principle by most serious writers on the international liquidity problem. But that logic assumes—and in fact insists—that the central bankers can be brought to accept the economic theorists' conclusion that the efficiency of the international monetary system requires the substitution for gold of a controlled international credit money, in the genuine sense of implementing it rather than merely paying lip service to it while actually believing the opposite. If that assumption is wrong—and it is not self-evidently right—then private speculation on a rise in the price of gold is more rational than academics and officials have been maintaining, and the refusal of the latter to consider the possibility of solving the international liquidity problem by raising the price of gold—a solution after all provided for in the Articles of Agreement of the IMF—may well turn out in the light of future developments to have been a Canute-like exercise in impotent vanity.

I begin my assessment of the present situation with respect to gold with some comments on the general nature of the gold standard. It is often asserted that the gold standard derives its attractiveness from the fact that its base money is a commodity esteemed for its own sake. This is the reverse of the economic truth: in a growing international economy, the provision of the additional gold reserves necessary to support expanding trade and payments at constant prices requires that the monetary price of gold be set high enough to generate an excess of supply over private demand, so that the private value is determined by the monetary value rather than the other way around. What really makes the system attractive to participating national monetary authorities is that, in contrast to any type of credit money system, gold gives them autonomy and protects their national sovereignty, by virtue of its character as an outside money, and that the quantity of it is determined by forces so complex that they can be regarded as outside anyone's control.

In this connection, theorists of the gold standard have placed great stress on the point that the system in reality contains an automatic control over supplies of new gold of a type conducive to economic stability, through the dependence of the profitability of gold mining on the relation between the price of gold and prices in general which determine costs of production. Inflation throttles and deflation stimulates new gold production and so govern the rate of monetary expansion in a price-stabilizing fashion. Critics have had no difficulty in demonstrating the practical defects of a stabilization mechanism that bases the money supply on a depleting mineral resource, subject to the vagaries of new discoveries and improved extractive technologies, and so durable that the existing stock is large in relation to current production and much more so in relation to price-cost-inducible variations in current production. A more fundamental criticism is that this analysis of the self-stabilizing properties of a gold standard ignores the fact that the use of a commodity for money is inherently inefficient, since the distinguishing characteristic of money is general acceptability and not the inherent qualities of the money substance. This inefficiency creates incentives for the substitution of credit moneys for the commodity money; and that substitution in turn permits inflation which gradually reduces the net flow of newly produced gold into monetary stocks to zero and turns it into a net outflow from monetary stocks into private use. Inflation on the one hand reduces the profitability of mining gold from existing mines and, more important, the incentive to prospect for new sources of the mineral. On the other hand it means that gold becomes progressively cheaper relative to other commodities to the private user, thus adding to the effects on demand of general economic growth a substitution effect on industrial and artistic demand and what might be termed a real balance effect on nonspeculative hoarding demand (i.e., hoarding not motivated by the expectation of a proximate increase in the monetary price)—it takes more and more gold, physically, to provide the same reserve of purchasing power. In consequence, inflation gradually converts the monetary authorities from the position of supporting the price of gold above the private market equilibrium price to that of suppressing the price below the private market equilibrium price.

From a logical point of view which recognizes the many inefficiencies of a commodity money system, this transition poses no serious problem: the monetary authorities should in logic be pleased enough to get rid of a non-interest-bearing asset whose real value is depreciating, and to replace it by a more efficient credit substitute. In reality, however, there is likely to be a sharp discontinuity in the thinking of the authorities. It is one thing to have to devise substitutes for gold in a system in which monetary gold stocks are increasing less rapidly than desired and quite another to have to surrender gold to the private market while willingly replacing it with what are thought to be inferior substitutes. And in view of this discontinuity, private speculators will be increasingly alert, as the net flow of gold into monetary stocks dwindles, to the possibility that the monetary authorities will choose to restore the inflow by raising the price. Further, speculation engendered by this expectation will reduce the monetary gold inflow and accentuate the necessity and difficulty of finding credit substitutes for gold; and this in turn will make it seem attractive to hold gold bought on a short-term expectation of a price increase on a revised expectation of a longer-term price increase. When the authorities are actually confronted with the need to sell gold to hold down the price, there will be disbelief in their willingness to do so, which will prompt further speculation on a price rise; and when in the face of this speculation they decide to discontinue gold sales, the speculators on a near-term price increase are temporarily confounded but have even more reason to hold onto their speculative gold purchases in the expectation of a longer-term price increase.

The facts of recent experience with gold fit this general theoretical analysis. Purchase of gold for private use was rising rapidly in the period up to immediately before the speculation associated with the devaluation of sterling. The upward movement took the form of a series of steps, in each of which the level of private purchases jumped and then stabilized at a higher level, consistently with the theory of conversion from short-run to long-run speculation just outlined. The Suez crisis boosted the annual average level from $350 million to $600 million; the rise in the free market price in the autumn of 1960, which led to the formation of the gold pool, raised it again to the $1,000 million mark; the Cuban crisis of 1962 boosted it to $1,100 million; in 1965, in consequence mainly of France's declared intent to take her surpluses thereafter in gold, the level rose to the $1,650 mark. On the supply side, gold availability from previously existing mines was falling steadily through the period from the early 1950's on; the rise in private demand was satisfied consistently with net additions to monetary gold stocks of $600 million on average from 1953 to 1964 by the growth of output from new mines in the Orange Free State on the one hand and growing Russian gold sales on the other. Production from the new mines had to peak eventually; and Russian gold sales are known to be uneconomic given production costs. In 1966 Russian sales stopped, and total new gold production peaked out. In that year, the free world's monetary authorities were converted from gold price supporters to gold price suppressors. To that change of basic position was added the short-term speculation on a rise in the gold price that began in October, 1967, and eventually forced the termination of the gold pool in March, 1968.

The basic question about the gold crisis of 1968 and the decision to end the single-market gold price system is whether, as American policy has assumed and a number of American international monetary experts believe, the system is still in the phase in which "normal" private industrial and artistic and long-term hoarding demands for gold fall short of normal supplies—including normal Russian supplies—so that leaving the private market to find its own level will eventually result in an equilibrium private market price below the monetary price; or whether a free market price above the monetary price is viable and sustainable. The answer to this question has been complicated by the fact that the Union of South Africa has become the source of three-quarters of new gold production, and, in the absence of Russian gold sales, has both an incentive to act as a discriminatory monopolist, by exploiting the IMF obligation to buy gold from members (so far as it can), and the capacity to hold off sales, so that the question of whether private demand would be willing to absorb private supply at a free market price above the monetary price remains to be tested.

The answer to this question depends ultimately on whether the central banks collectively will be willing to contemplate the replacement of gold by international credit reserves and to demonstrate this willingness effectively to the speculators by taking the appropriate steps to demonetize gold as a long-run policy. At present, they are in the weak position of having declared their belief that they have sufficient gold to satisfy monetary needs at the $35 an ounce price, but of having refused to back up that declaration by undertaking to trade in gold at that price with the free market. As long as they assert their belief in the validity of a price of gold which is below the private market's evaluation of the worth of gold but refuse to confirm the legitimacy of their belief by being willing to sell gold to anyone who disagrees with their judgment, they will rightly be regarded as hypocrites by the private market; and this will help to persuade the private market that they will eventually be forced to raise the gold price.

Are the world's central banks really prepared to contemplate the liquidation of the gold standard, or do they hanker after its restoration? The analysis thus far has explained why a gold standard system inevitably arrives at the point of having to choose whether to revitalize the system by raising the price of gold, or to replace it by the substitution for gold of international credit money based on international monetary cooperation. The world economy has arrived at this juncture before: once, in the 1890's, it was saved from having to make the choice by new gold discoveries that provided the required increase in monetary reserves; once, in the 1930's, it reached the limits of international monetary cooperation and had to settle, willy-nilly, for a rise in the price of gold effected by a painful process of cumulative devaluation, incidentally highly disruptive of world trade. What signs, if any, are there that in present circumstances the world's monetary authorities are prepared deliberately to choose to substitute credit money for gold in order to circumvent a rise in the monetary price of gold?

The signs really cannot be said to be promising. Much can indeed be made of the degree of international monetary cooperation that has developed over the past decade. But that cooperation has been concerned with economizing on the use of a gold supply that has been growing less rapidly than has been desired, not with explicit endeavors to demonetize gold in favor of an international credit reserve. On the face of the matter, there is every reason to expect nations to cringe in the face of the realities of gold demonetization. The trend has been towards closer international economic integration and interdependence; but this has been in conflict with the demands of national sovereignty and independence. The survival of gold as the basic international reserve money has preserved an element of autonomy and independence, of capacity to claim to be cooperating in the solution of international problems while exercising pressure for the preservation of national self-determination and of national influence on world affairs. The anonymity of gold serves national interests still considered vital; and the capacity to defend national interests considered vital will not readily be sacrificed, whatever the theoretical arguments and the general world interest in doing so.

The hope of those who look to replacement of gold by a more rational system based on international credit money is founded on the agreement to initiate SDR's and the prospects of their developing into a substitute for gold. The question therefore arises whether SDR's could constitute an international money useful enough eventually to substitute for gold. Much of the criticism of the SDR scheme has focused on the changes in voting rights in the Fund agreed as the price of acceptance of it, and their adverse implications for the expansiveness of future international liquidity policy. A far more important point for criticism is the acceptance of a gold guarantee on SDR's. This guarantee means that SDR's will be less useful to potential deficit countries than other forms of credit finance, and as useful as gold to potential surplus countries, so long as there is any possibility of a rise in the monetary price of gold, and this depends on the attitude of the central banks involved. So long as there remains an expectation among central bankers that the monetary price of gold may have to be raised, the use of SDR's will tend to be confined to deficit countries that would otherwise have to settle in gold—and that means the United States, which would thereby achieve only a temporary and purely cosmetic improvement of its gold reserve statistics.

It therefore appears that, contrary to the logical arguments according to which the authorities by their action in March have taken a firm step towards demonetizing gold and replacing it by international credit money supplied by some sort of world central banking, in fact the authorities have succeeded only in nominally carrying international monetary cooperation in crises a step further than has been reached in previous confrontations of the problem of the gold standard, while still leaving open the question of the monetary price of gold, and indeed increasing rather than decreasing the likely necessity of an eventual rise in the gold price.

What are the options? I would myself much prefer the theoretically logical solution, replacement of gold by international credit money, though along with that solution I would like to see much more flexibility of exchange rates than the present system entails. But if, as I strongly suspect, the central banks are for reasons I have given not willing to take that step, the international monetary system will become increasingly evidently an absurd arrangement, a charade in which gold is traded at a monetary price below the world market price among official transactors who will be aware that the official price of gold increasingly underrepresents its scarcity for monetary purposes. There will be an increasing danger of a scramble for gold among central banks, gold which must come from the United States, and other nations will increasingly be able to blackmail the United States, using gold as the threat, into extending policies of balance-of-payments intervention—especially with respect to private capital exports—which have already gone far towards subverting its free competitive system. The United States might be goaded into, or might choose, to forestall these pressures by declaring the dollar inconvertible into gold by other monetary authorities, as Britain did with sterling in 1931. But the results of so doing would probably again be a period of exchange rate flexibility, international monetary disturbance, and increasing restriction of trade and payments, ending in a return to fixed rates pegged on gold at a higher price. It was a repetition of the 1930's experience of raising the price of gold the hard way that the postwar planners of international monetary reform sought to forestall by providing explicitly for the possibility of raising the price of gold by agreement of the members of the International Monetary Fund. In view of the prospective dangers that the present two-tier gold price system is courting, the time may have come for international monetary experts and officials to consider seriously the prospect of activating that provision. We all know the arguments against it, and we are all professionally prejudiced in favor of less irrational solutions; but we should perhaps swallow our intellectual pride in face of the prospective dangers of the present gold situation and turn our attention to devising solutions to the manifold problems that a rise in the monetary price of gold would entail.

DISCUSSION

Edward M. Bernstein: We have had three very interesting papers bearing on the future of gold. I regret that I have not had the benefit of Professor Mundell's paper on real and paper gold. I have read his theoretical analysis relating the rate of gold production, the growth of output, and monetary expansion to prices and interest rates in a gold standard world. Mundell has ingeniously adapted his analysis to a gold exchange standard. As he shows, with gold production rising very slowly and actually less until 1962 than it had been in 1940, the international monetary system became increasingly dependent on dollars for reserves. Because the substitutability of dollars for gold is not unlimited in reserves, the gold exchange standard became more difficult to sustain as gold became scarcer and dollars more plentiful. I am not sure whether Mundell believes that a similar problem would arise if a fiduciary reserve asset created by the IMF were issued in the amounts necessary to provide for an adequate growth of aggregate monetary reserves. The critical question regarding the future role of gold is how to make the international monetary system function with multiple-reserve assets—gold, dollars, and Special Drawing Rights.

Professor Machlup has given us an enlightening discussion in his "Speculations on Gold Speculation." He may be right that in the broad sense in which he uses the term the private absorption of gold has been mainly by speculators, including traditional hoarders. It may be that speculators will unload their enormous accumulations of recent years as soon as they are convinced that the U.S. Treasury will not raise the monetary price of gold. And it may be that even traditional hoarders will again be large sellers of gold, as they were from 1931 to 1938, if there is a world depression, or the price of gold shows indications of falling sharply, or if other investment assets become more acceptable to them.

It is not relevant to the question of the future role of gold to speculate on what the price in private markets may be five or ten years from now. Long before then, the monetary authorities will have taken irrevocable decisions on the monetary role of gold. Insofar as the price of gold in private markets can affect the views of central bankers on the issues of international monetary reform, it will be the price during the next two or three years. I have expressed the view that the price during this period may be expected to fluctuate between $38.00 and $42.00 an ounce, falling and rising with changes in confidence in the dollar and other currencies. I do not rule out the possibility that if the dollar should become much weaker, the price might rise for a time to $45.00 an ounce; and if the dollar should become much stronger, speculators might unload their gold holdings and the price could fall to $35.00 an ounce or less for a time.

I think South Africa's desire to sell some gold to the monetary authorities at $35.00 an ounce is not for the purpose of exploiting its quasi-monopoly position as the producer of three-fourths of the gold outside the Communist countries. I believe that the real purposes of South Africa are to establish the eligibility of newly mined gold for incorporation in the monetary stock; so that if there is a uniform change in the monetary price of gold, it will benefit from it by selling its output to the monetary authorities at the higher price. To a lesser extent, I believe that South Africa wants to assure itself against the loss that might result from a sudden unloading of speculative holdings if this should occur and drive the price to $35.00 an ounce or less.

There is an exaggerated idea of the extent to which South Africa can act as a discriminatory monopolist—selling some gold at $35.00 an ounce to the monetary authorities to force up the price in private markets, say, to $40.00 or $45.00 an ounce. This is not a practical policy for South Africa. First, that country accounts for only three-fourths of the output in the Western world and a high price might bring forth larger supplies, not only from other producers, but from the Soviet Union. Second, as Machlup has shown, the supply of gold is not confined to new production, and high prices in the private market could call forth large sales by private holders. Third, one side of the discriminatory price system is fixed—the monetary price—and for every ounce of gold sold to the monetary authorities, South Africa would have to surrender \$5.00 or more an ounce in order to reduce the private supply. South Africa does not have a sufficiently complete monopoly, and the demand for gold in private markets is not sufficiently inelastic, to enable that country to profit from a diversion of supply from the private market to the monetary authorities, if the private price is \$40.00 or more an ounce.

If I am right, the monetary authorities will be unable to compete with speculators for the newly mined gold at the monetary price of \$35.00 an ounce. In a speech in New York on January 23, 1968, before the gold crisis that led to the termination of the gold pool, I said: "The noteworthy aspect of the gold speculation is not that it was so great, but that it was not much greater. In a world in which the [world] money supply is measured in the hundreds of billions of dollars, in which foreign holdings in the United States are counted in the tens of billions of dollars, speculation which absorbs about \$1.5 billion of monetary gold is really miniscule." The speculators proved in March, 1968, after they had acquired about \$3 billion of gold at \$35.00 an ounce, that the monetary authorities are not prepared to sell gold on the massive scale and for the indefinite period that would have been necessary to quench the speculative demand for gold in a world in which confidence in currencies had been seriously undermined. When the gold pool said the existing stock of monetary gold is sufficient in view of the prospective establishment of the facility for Special Drawing Rights, they did not mean that they were prepared to see a decline in the monetary gold stock, but that they were reconciled to no further growth of total gold reserves.

Professor Johnson in discussing the "Gold Rush of 1968 in Retrospect and Prospect" is undoubtedly right that the world's central bankers are not really prepared to contemplate the liquidation of the gold standard. I do not think that this was in their minds at all when they issued their communique on March 17, 1968. But Johnson has not exhausted the logical alternatives. It is not a question of either retaining the gold standard in its present form by a uniform rise in the price of gold or the replacement of gold by international credit reserves. There is a third alternative, and that is to retain the present volume of gold reserves in the international monetary system and to provide for the growth of reserves in the future through a fiduciary reserve asset created by the IMF, in amounts necessary for the trend growth of the world economy, without regard to the balance-of-payments needs of any country. As I understand this third alternative, monetary reserves would consist of a fixed amount of gold (now about \$40 billion), a fixed amount of foreign exchange reserves, mainly dollars (about \$25 billion), and a growing amount of Special Drawing Rights.

In a system with multiple-reserve assets, the preference of central bankers for gold could undermine the international monetary system; and this preference is not likely to become smaller in the future if the price in private markets remains above \$35.00 an ounce. I have long left that if the gold standard is to survive in any form it will have to be as a composite gold standard—a term I first used in 1962—under which different reserve assets would be used together in international reserve transactions. My proposal is to have all countries earmark their gold and other reserve assets in a Reserve Settlement Account in the IMF. They would thus retain title to the precise reserve assets they hold. But in using reserves, deficit countries would draw down their balance in the Reserve Settlement Account, reducing their holdings of the different reserve assets pro rata. And surplus countries, in acquiring reserves, would build up their balance in the Reserve Settlement Account, acquiring the different reserve assets in the average ratios in which they are used by the deficit countries.

I cannot see much merit in a massive increase in the price of gold accompanied by the hope that this will result in an appropriate growth of aggregate monetary reserves in the future. The composite gold standard seems to me not only a more rational but a more practical proposal for dealing with the gold and reserve problem. This is a natural evolution of the gold standard and I am confident that central bankers will gradually realize and accept it. In the past twenty-five years there has been an encouraging acceptance of the basic principle that the international monetary system must evolve to meet changing conditions. This farsighted attitude has made possible the creation of the International Monetary Fund at Bretton Woods; the increase of its resources from $7 billion to $21 billion; the formation of the Group of Ten with a commitment to provide $6 billion of resources to the IMF when this is necessary to meet a threat to the international monetary system; the increase of reserve credit on a massive scale through reciprocal currency arrangements (swaps); and the adoption of a new fiduciary reserve asset (Special Drawing Rights), a landmark in international monetary evolution.

I think I am a reasonably good judge of what is practical in the way of international monetary cooperation. In my opinion, a composite gold standard, consisting of gold and other reserve assets, will be adopted in the course of the next few years. It may be operated through a Reserve Settlement Account or through some other institutional device. I think that central bankers recognize that the only way in which the gold standard can survive is through further evolution. The composite gold standard provides an appropriate role for gold as the basis for currency parities, as the principal component of monetary reserves, and as the link to the equivalence in value of all reserve assets.

FRED HIRSCH: Economists like to count it among their professional responsibilities to tell bankers and economic administrators what it is that they, the practical men, actually do. This is as it should be. But it is not obvious that economists enjoy similar professional competence in telling the practical men what it is they actually think. To put it differently, the comparative advantage of economists may not be especially high in the field of parlor psychology, whether applied to central bankers, governments, or other collective groups. I do not intend, therefore, to cross swords with Harry Johnson on the psychological and behavioral issues that underlie his paper; that is to say, on what the true feelings of central bankers and other relevant decision-makers about gold as against reserve credit assets actually are and how far these feelings are what the financial officials say they are. Rather, and more dangerously for myself, I want to dispute some of his economics or, more exactly, political economy.

I shall develop three main points. The first is to stress that the characteristics of a gold standard are vitally affected by whether or not the system is managed in the sense that discrete changes are made in the international gold price. An international gold standard may have advantages of anonymity and of diffusion of decision taking or it may have advantages of administered regulation, but it cannot have both advantages together. My second point is to stress that monetary absorption of gold and nonmonetary demand are to a significant degree interdependent; so that whether at any time the monetary authorities are supporting or suppressing the gold price (in relation to the private market equilibrium price) may itself be dependent on the current and expected policy of the monetary authorities in the matter of gold absorption. I then draw on the implications of the first two points for the so-called "autonomy" conferred by gold as a reserve asset.

It has become fashionable again to point up the automatic control element, or more properly the self-equilibrating element, in the gold standard, which Keynes noted in the *General Theory,* deriving from the interplay of changes in price levels on gold supply and demand. This revival is somewhat curious, since there are reasons to believe that the self-correcting forces have become less strong in recent times, at least on the production side. Gold output has become less responsive to changes in the price level, for reasons I have developed elsewhere.[1]

This aside, it is important to remember that the automatic control, insofar as it operates, does so through a crisis system. Thus, an expansion of gold output or, following Harry Johnson, a substitution of credit money for gold money, permits inflation; this reduces or eliminates the flow of gold into monetary stocks, which in turn threatens deflation or collapse of the system or (as in the 1930's) both. Note that even with the aid of deflation, a discrete, administered increase in the gold price (i.e., supplementation of the automatic forces) is needed for the equilibrium of the system to be restored. Hopefully, such an increase, undertaken in time, may avoid the deflation as well as the collapse. But if this is done, such increases in advance of crisis unquestionably imply a managed system.

---

[1] "Influences on Gold Production," IMF *Staff Papers,* Nov., 1968.

The question then is: under such management, how many of the attributes commonly associated with the international gold standard remain? Remember that the effective price decision is taken—and no one suggests that this is no longer true even today—by one country, the United States. Moreover, within that country, the final say is not in the gift of the bodies that are most susceptible to international influence, the Federal Reserve or the Treasury or even the President. The power of ultimate decision is vested in the U.S. Congress, an institution not hitherto penetrated even by the Group of Ten.

The implications of this for the role of gold in its anonymous function as an international asset are of course not limited to the occasion of a price change itself, since the possibility of such a change—or the potential refusal to make such a change—are contingencies hanging over the value of countries' gold reserves at all times. Thus, while the quantity of gold may still, to quote Harry Johnson, "be determined by forces so complex that they can be regarded as outside anyone's control," this is no longer true—if the system is to be safeguarded against collapse—of the value of gold reserves, which is the relevant entity. Indeed, the value of future gold reserves is almost certainly more closely in the control of a single country than the future value of international credit reserves in their present or prospective forms. In these circumstances, can gold in its reserve function really be regarded as "outside money"—or is it nearer to the fiduciary issue of the United States?

Not only can the monetary authorities determine the price of gold by policy decision; they are also free to determine how far they should hold gold rather than other instruments as reserve assets, and beyond this, they can in principle regulate their need for reserve assets as a whole. If private demand for gold, in the same way as private demand for most commodities, including silver, were predominantly a commodity demand in the sense that the commodity was "used up" or retained primarily for the psychic income it could provide, then the impact of different levels of present and future official demand on the equilibrium of the gold market would be relatively straightforward. In practive, however, "consumption" of gold in the narrow sense has recently accounted for a minor portion of total private use, as Fritz Machlup has discussed. In the bulk of such use the primary demand is not for gold, but for a gold claim; or more exactly, for the (presumably discounted) differential between the expected future value and the present value. Whether gold held for such purposes is categorized as speculation, savings, hoarding, or investment depends in part on taste and in part on one's time horizon. The distinction between short-term holdings and longer-term holdings is a subsidiary categorization between different financial investment assets, demand for which is always governed by expected future value in relation to current value.

It follows that the impact of a change in the current price alone predicts little about the effect on this sector of private demand, geared as it is to the differential between the current price and expected future price. As long as anticipation of offical policies relating to the use of gold as a reserve asset continues to influence "investment" demand, actual current absorption of gold into monetary stocks is not a sufficient test of official influence on the gold market. Specifically, absence of such monetary absorption at a time of uncertainty as to future official demand does not indicate, *pace* Harry Johnson, that the monetary authorities are suppressing the gold price below the private market equilibrium price. Private gold demand is a partly dependent variable of official gold absorption and official policies on gold stocks and not an independent component of total demand. To refer to the present official price of gold as a disequilibrium price, as other critics have done, is therefore to beg the question of what these future policies on gold reserves will be. Of course, once private demand of a "noninvestment" nature itself absorbs the whole of current supplies, including such supplies as may become available from stocks, then anticipations of future official absorption policies will no longer play the same role.

If both the official valuation of gold and the volume of private demand are dependent on particular national decisions, then the attributes of gold as a private monetary asset and especially as a reserve asset will reflect this susceptibility to national management—and mismanagement—in a number of ways. The point is far from a new one: it was at the center of Keynes's assault in his *Tract on Monetary Reform.* One may recall Keynes's patriotic warning against making Britain's finances dependent on the policy actions of that untried and doubtfully responsible body, the Federal Reserve Board of the United States.

The experience of the past half century does not suggest that Keynes's emphasis on the role of policy decisions in this sphere was excessive. Indeed, in the light of this experience, it is rather remarkable that the attachment of attributes such as anonymity, security, and sovereignty to possession of gold reserves rather than other reserve assets has not received closer examination. As a private asset, gold continues to possess the first two of these attributes in a high degree. Their extension to gold holdings of a monetary authority is a questionable exercise in composition. Moreover, little evidence can be found to suggest that a belief in such attributes as anonymity or sovereignty—as distinct from the authorities' assessment of possible loss or gain from valuation changes vis-à-vis reserve assets lacking a gold guarantee—has influenced the behavior of monetary authorities. The known financial history of World War II shows a number of cases of uncertainty attaching to the usability of gold reserves, both as to acceptability and to hazards of transportation and looting; if anything, the period saw a shift in reserve asset preference toward foreign exchange.

Nor, observing the experience of countries in recent years, is it easy to discern how "the survival of gold as the basic international reserve money has preserved an element of autonomy and independence" for national governments. Aside from the central issue of influence over the exchange rate, the key aspects of monetary sovereignty appear to be the level of aggregate reserves and the ability to supplement reserves by calling on credits without unwanted conditions. For France in recent months, the latter element—the elasticity of swaps—would seem to have been decisive in permitting national decisions on other aspects of economic management to be maintained; whatever the significance of the composition of reserves between gold and other assets may be, it has not been evident here. All this suggests that if certain attributes of gold, such as a firm value guarantee, are duplicated in instruments such as Special Drawing Rights, these instruments may not be less desired vis-à-vis gold simply because they are monetary claims without being monetary goods.

These remarks should not be misunderstood. There are important differences between the attributes of gold and other reserve assets, residing mostly in the physical properties of gold. From these different attributes, as well as from custom, decision-makers in central banks and governments may have different attitudes to gold than to other reserves. They may take different views on decisions to supplement reserves by means of attaching a higher valuation to gold metal than by means of making use of the special instruments they have collectively set up for the purpose of reserve supplementation. On these differences of attitudes, in my assessment, economists can have less to say than Harry Johnson has suggested. The evolution of gold management and of economic management at large has diminished the economic distinctions and also the political distinctions between a number of aspects of gold and nongold liquidity.

MIROSLAV A. KRIZ: What would a commercial bank economist be were he not a realist? From this vantage point, I shall deal with gold as it is—in a politically, institutionally, and culturally diverse world, which is monetarily unsettled, insecure, and suspicious. I shall not discuss what might become of gold, and all that, in a politically, institutionally, and culturally unified and monetarily secure world if it ever comes.

In the world as it is, governments and central banks are not ready to replace gold with U.S. dollars or money controlled by an international board. This was once again evidenced by the decision, in the midst of the gold crisis in March, 1968, to stop supporting the price of gold in the London market although the cost of intervention equaled only 10 percent of the total of the gold stocks of the seven gold pool participants. It was also evidenced by the fact that, in the midst of the crisis, the seven countries represented at the Washington Conference chose, among all the options open in the circumstances, the smallest possible change in the world's gold arrangements—not a demonetization of gold but a hybrid in the guise of the so-called "two-tier gold price." The continuing demand for gold was also evidenced by the fact that, beginning shortly after the Washington Conference, the opportunity to add to official gold stocks out of the substantial amounts sold by France and the International Monetary Fund was seized by a strikingly large number of countries.

Evidently, governments and central banks—to paraphrase Professor Johnson's words—do not accept the economic theorist's thought that an international monetary system in which gold would be replaced by controlled international money would be a more efficient institutional arrangement than gold. Professor Johnson regards the preference for gold as irrational. He holds it against the central bankers that they are merely paying lip service to the substitution of controlled international money for gold while actually believing the opposite.

Yet, to my mind, given all the circumstances of the real world, there may well be deep-seated rationality in gold. The rationality has, of course, nothing to do with any mystical properties of gold. It rests on the qualities of gold as "outside money"—money independent of any government or group of governments and, therefore, more suitable as international money than a national reserve currency like the dollar or money controlled by an international board.

Of course, the two-tier price system—as I have already implied—is not rational, for it is neither a two-tier price nor a system. It is not a two-tier price because the official $35 parity is an act of the U.S. government embodied in law, not a market price. It is not a system since the world has merely returned to the state of affairs that existed from the end of World War II to 1961—a state of affairs that was justifiably regarded as unsystematic and was, for that very reason, changed beginning in 1961 when governments began to intervene in the London gold market to support the $35 price.

This recent illogicality must not, however, be allowed to overshadow gold's intrinsic qualities that, in the world today, cannot be replaced, with the same essential advantage, by any other national or international arrangement.

First, gold is beyond the control of any one nation, especially as redistributed today, with the United States holding only slightly more than one-fourth of the world's monetary stock. As a result, gold makes it easier than any alternative arrangement—whether an inconvertible dollar standard controlled by the United States or money controlled by an international board—for governments and central banks to safeguard a substantial measure of autonomy and independence in their own monetary and fiscal matters. For acceptance of an inconvertible dollar exchange standard or of money controlled by an international board would be like having two central banks—one at home and the other outside the country—sometimes working at cross-purposes.

The fact that gold is beyond the control of any one nation is also a quality that has not remained unnoticed by continental European central bankers. In their view, the experience in the United States over the past decade points unmistakably to the conclusion that the U.S. monetary authorities and the American people are impressed by U.S. gold losses and concern over the nation's gold stock and not by increases in U.S. dollar liabilities to foreign governments and central banks.

Another unique quality of gold is that there are times and circumstances where no other money—whether dollars or money controlled by an international board—will do, because gold alone is universally acceptable, without any questions being asked, as a means of payment of last resort. However remote it may appear, a contingency that dollars or international money might be blocked cannot be excluded in times of political tension.

Yet another quality of gold is that its value as a common denominator of currencies is not necessarily immutable. It is, of course, for this very reason that those in most of the large industrial countries responsible for administering monetary reserves are holding in gold 60 to 90 percent of their nation's total gold and dollar reserves. Furthermore, and quite significantly, they have succeeded in endowing two of the most recent international monetary innovations—the General Arrangements to Borrow and the Special Drawing Rights—with absolute gold clauses. By absolute, I mean that, unlike the IMF gold clause, the GAB and SDR clauses cannot be rescinded by the IMF Board of Directors in the case of a uniform rise in the price of gold.

Considerations like these are dismissed by some students of international liquidity by paraphrasing the reply that Maurice Chevalier made when asked how he felt to be eighty: "Very nice—considering the alternative." The alternative is, of course, gold demonetization—cutting loose from the ancient fetish. This school of thought regards the convertibility of dollars into gold as a fiction maintained to reassure continental Europe and to comply with certain technicalities of the IMF charter. It is, of course, true that convertibility becomes nominal when governments and central banks have little choice but to refrain—for nonmonetary reasons—from exercising it. But it is illusory to expect that a formal gold embargo by the United States would somehow lead to a worldwide demonetization of gold. Amidst uncertainties about currency values, not only would international monetary cooperation be severely damaged, but nations would insist on settlements in gold, as in the 1930's. Paradoxically, but quite rationally, the consequence of a U.S. embargo would be to raise gold to unchallenged primacy in international settlements—to the detriment of the dollar and international credit in all its forms, including intercentral bank and intergovernment credits, the IMF, and, hopefully, the SDR's.

Since governments are unwilling to give up gold, its price must, obviously, be appropriate to enable gold to function properly as the keystone of the international monetary system. The argument is economic. It is also political. Having disagreed with Professor Johnson about the rationality of gold, I do wholeheartedly agree with his pragmatic conclusion that "we should perhaps swallow our intellectual pride in the face of the prospective dangers of the present gold situation, and turn our attention to devising solutions to the manifold problems that a rise in the monetary price of gold would entail."

This brings me to Professor Machlup's helpful—and elegant—contribution to the psychology of gold speculation. To my mind, however, he does not do justice to the rationality of gold speculation.

The Machlupian analysis basically runs in terms of gold versus the dollar; and, on this basis, it is true, those foreigners who can, legally or illegally, worship the cult of American equities—and, obviously, the Americans themselves—have done better than if they had worshipped the golden calf, a cult which is illegal for Americans. They have done better because, in the rich, productive, and flexible U.S. economy, purchases of suitably selected stocks offer real promise of capital gains and, hence, protection against inflation. Today, with a 4½ percent rate of inflation that, if maintained, would in sixteen years reduce the buying power of the dollar by one-half, Americans are also running for cover, on a vast scale, by buying land, diamonds, old masters, antiques, and even vintage motor cars. They have learned to live with the depreciating dollar without buying gold. But speculation about gold speculation in dollar terms would become a less speculative exercise if there were a free gold market in the United States—a market open to everybody, not just to holders of Treasury licenses, as today.

However, for most of those who buy gold as "safe investment"—to use the Machlupian terminology—what matters is the increase in gold price not in terms of the U.S. dollar but in terms of their own currencies. A Frenchman, who has lived through a dozen devaluations since 1914 as a result of which the franc has lost more than 99 percent of its value, has succeeded rather well in protecting that part of his patrimony that he has kept in gold: Having purchased a one-kilogram bar in 1914 for 3,100 francs, he can sell it today for 700,000 "old" francs, equivalent to 7,000 "new" francs. The buying of gold has similarly—though to degrees varying according to country—provided protection to holders of German marks, Italian lire, Belgian francs, Argentine and Chilean currencies, etc. The search for protection does not stem from irrational speculation but from deep-seated causes that must be taken very seriously. For speculation can never get a grip on a currency managed under a healthy and clear policy. If a currency is threatened, it is because people are uneasily aware that the longer they hold money, the less they get for it.

Altogether apart from the precautionary buying of gold, there is the hedging against the consequences for exchange markets and exchange rates of a U.S. gold embargo and the accompanying international monetary disturbances. That an occurrence of this sort cannot be dismissed stands out clearly from Professor Johnson's paper. Corporate treasurers and others who administer large amounts of cash are continuously compelled to take a view of future values of currencies, for these will determine the future amount of their profits or losses. Speculation in this sense is a normal precautionary activity—given all the circumstances in the world today, including three monetary crises in twelve months.

To conclude, the manifold efforts over the past few years to safeguard and improve the international monetary structure and, with it, the meaningful convertibility of currencies and basically free, multilateral and nondiscriminatory trade and payments cannot be brought to a successful conclusion unless gold itself is dealt with as a fundamental subject in this field. With the international monetary structure unchanged in 1968 except for the divorce of the monetary from the market gold price, the great unfinished business remains intellectually as provocative, and monetarily and politically as challenging, as ever. The most essential task today is to face the matters governing gold in an emotion-free atmosphere so as to avoid miscomprehension, mistrust, and mishandling.

ROUND TABLE ON EXCHANGE RATE POLICY

FRITZ MACHLUP, *Chairman*

GOTTFRIED HABERLER: What I have to say is, I am afraid, not particularly new or striking. In fact, I must confess that my thinking on exchange rate policy is still based on what our good and highly esteemed friend Robert Mundell recently called "the primitive stage in which the argument for exchange flexibility was put by Milton Friedman, James Meade and Friedrich Lutz." But economic policy is too serious a business to justify sacrificing any wisdom or usefulness there may be to a frantic search for novelty, unnecessary jargon, or elegance.

The system of fixed par values of currencies is still the law of the monetary globe except for highly inflationary countries where it would quickly lead to intolerable consequences. Under this system, changes of exchange rates are few and infrequent and each change becomes a major operation with serious economic and political repercussions.

The rules of the game, which must be obeyed to make the system of fixed exchanges work smoothly and efficiently, are well known. Deficit countries must restrain monetary expansion by not obstructing automatic forces which tend to bring it about or by positive action. Surplus countries must stimulate monetary expansion by not counteracting expansionary automatic forces or by positive action. I take it for granted that it is undesirable to try to suppress deficits by specific controls. I also assert that controls, if they are to have an appreciable and lasting effect, must be comprehensive and since they tend to become leaky as time goes on must be continuously tightened and expanded.

However, I sharply distinguish a policy of suppressing deficits by controls from a policy of financing them by *ad hoc* borrowing in one form or another. I mention this, to my mind, very important distinction because Harry Johnson in a recent paper lumps the two types of policies together under the heading of "financing" (in the broad sense in contrast to real adjustment).[1]

The postulates of balance-of-payments adjustment under fixed exchanges have increasingly come into conflict with domestic policy objectives—full employment, growth, and price stability. Moreover, in all countries it has become more and more difficult to reconcile the objective of price stability with that of full employment and growth. Changing the fiscal monetary mix cannot resolve the conflict between external and internal balance except perhaps temporarily. Moreover, if it could, it would open a conflict between employment and growth.

From the point of view of external balance, the important fact is that different countries have different priorities and preferences as between different objectives, and that their record of success or failure in reconciling conflicting policy objectives has been very unequal indeed. Some have a comparatively greater aversion to unemployment and others to inflation. Whatever is true in the long run, there certainly exists a short-run trade-off between price stability and employment. But the trade-off probably changes from time to time and there is no reason why it should be the same in different countries. Wage push exerted by labor unions is of different intensity in different countries. In some countries, say the United Kingdom and United States, trade unions are more aggressive and irresponsible than in others, say Germany and Japan.[2]

It stands to reason that these differences and disparities produce in different countries different trends in costs of production and prices, different growth rates in nominal and real income, and it is not surprising that in such an environment imbalances pile up frequently. These disequilibria require ever increasing *ad hoc* financing, induce countries to apply wasteful and distasteful specific controls and from time to time, when tension gets too strong, produce violent explosions in the form of acute balance-of-payments crises.

---

[1] "Problems of Balance of Payments Adjustment in the Modern World," *Essays in Money and Banking in Honour of R. S. Sayers* (London, 1968), p. 122.

[2] If there were no wage push, it would be much easier to achieve external and internal balance at the same time without inflationary bias for the world as a whole, even if wages are rigid downward. Deficit countries would simply keep the wage level stable (by monetary restraints) until prices and costs have been reduced sufficiently through the gradual increase in labor productivity, and surplus countries would let their wage level rise a little faster than productivity, thus allowing prices and costs to rise.

This is not a satisfactory state of affairs. In the framework of fixed exchanges, ruling out direct controls, the problem could be solved only by achieving a much greater harmonization and mutual adjustment of basic economic objectives and policies than we have at present. Setting up of an international central bank either full-fledged as recommended for years by Robert Triffin or incipient in the form of a reserve settlement account as recently proposed by Edward Bernstein, Fritz Machlup, and Robert Mundell would, by itself, not help unless such an institution was given broad power to coordinate monetary and fiscal policies, which under present conditions means setting priorities on major policy objectives—full employment, growth, price stability. It would, of course, always be possible for an international central bank to create enough money to accommodate the most inflationary member or, more generally, the member with the weakest balance of payments. (A sufficiently adverse shift or trend in international demand can put a country in serious deficit which has not engaged in inflationary policies. But I believe that such cases are very rare indeed.) This solution, however, implies policy harmonization at the most inflationary level and is not acceptable to the less inflationary countries.

What are the chances, speaking only of the industrial countries, that in the future sufficient harmonization of policies will be achieved? It is true there has been much talk about the necessity of policy coordination, and international and interregional discussions and negotiations to bring this about are going on almost around the clock in the IMF, OECD, EEC, etc.

Recent events have shown, however, that even among the comparatively homogeneous and closely knit group of countries organized in the European Common Market harmonization of basic economic policies has not progressed far enough to ensure smooth balance-of-payments adjustment. This is all the more serious because the EEC's highly interventionist, extremely costly, and shockingly wasteful agricultural policy makes exchange rate changes exceedingly difficult. (This stems from the fact that the high uniform prices guaranteed to farmers in all six countries are expressed in terms of gold dollars and member countries are obliged to reduce domestic currency prices paid to farmers in case a country's currency appreciates and to increase them in case a country's currency depreciates. Thus any change in exchange rates creates grave political and financial problems and the equilibrating balance-of-payments effects of a change in the exchange rate are sharply reduced by eliminating the farm sector from the adjustment process.)

If what I have said, especially about the crucial importance of international policy harmonization, is broadly correct—and frankly I cannot see how it could be denied—and if it is agreed that selective controls are to be avoided, then it follows that some flexibility of exchange rates is the only possible solution. Frequent discrete changes—in other words, adjustment of rigid pegs at short intervals—maximizes speculative disturbances. Continuously flexible rates are a much better solution. As far as the developed countries are concerned, I believe that limited flexibility, a wider band, and more or less automatic adjustment of the band itself would be sufficient and probably preferable to unlimited flexibility, at least at the beginning until policy-makers and the public at large have become familiar with the new system.

Consider for a moment how this system would have worked in 1968. Before the student and workers rebellion in France of May and June, the French franc was strong and its value in international markets would have been near the upper end of the band. Then it would have fallen to the lower end of the band resulting in an automatic devaluation of 8–10 percent depending on the width of the band. This would certainly have brought considerable relief over an extended period by improving the current balance.

In the worst case, speculation and inflation would have kept the rate at the lower end of the band. This would still have been an improvement over the actual situation because of the favorable effect of the automatic devaluation on the current balance.

If the band is adjustable, it would be lowered by small steps, say, one-sixth of one percent per month or at shorter intervals, in case the rate stays at the lower end of the band for some specified period.[3] If the policy of limited flexibility can be made credible, that is to say, if responsible monetary and fiscal policies are pursued so that it becomes generally accepted that no larger devaluations than the stipulated maximum will become necessary, no disturbing speculation need be expected. For the inducement to export capital in anticipation of a decline in the rate of 2 percent per year can be offset by a corresponding interest differential.

It is, of course, always possible that inflationary pressure becomes so strong or political disorders so serious that the system of limited flexibility collapses just as the system of the fixed exchanges breaks down under such circumstances. All that is claimed is that the system of limited flexibility can cope with much more serious tensions and disturbances than the present system of fixed exchanges.

Frankly, I find it exceedingly difficult to think of valid objections to flexible exchange rates. The popular argument that they would give rise to disturbing speculation may be true if comparison is made with a utopian gold standard where rates are really sacrosanct (except in case of war and revolution). But this state of affairs, whether desirable or not, is beyond reach. On the other hand, when compared with the adjustable peg system, flexible rates are undeniably much less prone to incite speculation.

The equally popular objection that fluctuating exchange rates hamper international trade and capital movements is subject to the same type of rebuttal. To be sure, stable rates are a great boon for international trade and payments, provided free convertibility is maintained and overriding domestic policy objectives need not be sacrificed to keep the rate nominally stable. If, in order to protect the fixed exchange rate, it becomes necessary to impose extensive controls and to tolerate more unemployment and slack than is required by domestic policy objectives, then the fixity of the rate becomes a sham and the ends are being sacrificed to the means.

Viner's classical argument that flexible exchange rates would undermine domestic financial and wage discipline I find unconvincing under modern conditions. Compared with a shrinking gold reserve, a persistent decline of the rate would provide an equally strong inducement to restrain inflation and an even clearer alarm signal to the monetary authorities that something is wrong.

Some writers seem to believe that the modern theory of optimum currency areas provides a new sophisticated argument against flexible exchange rates between existing countries on the ground that few, if any, existing currency areas are of the optimum size or shape.

Nobody would, of course, deny that small countries may be well advised to peg their currencies to that of some big neighbor or to form a fixed exchange rate bloc with other small countries, because the inconvenience and mechanical cost of fluctuating rates vary inversely with the size of a country.

But even a very small country, if surrounded by highly inflationary or deflationary neighbors, would find that a flexible exchange rate is the lesser evil. The same is true if a small country itself is subject to much greater internal monetary instability than its large neighbors. Switzerland in some periods of its history offers an example of the former case, Uruguay of the latter.

If balance-of-payments disequilibrium is due not to purely monetary causes (differential inflation), but to shifts in international demand (nonmonetary factors) or if monetary and nonmonetary factors are involved (the latter may be autonomous or caused by the former), the optimum currency area theory demonstrates that existing countries (currency areas) are not of optimum size or shape from the point of view of efficient balance-of-payments adjustment by exchange rate changes.[4] It follows that if political boundaries could be redrawn in a suitable manner, exchange flexibility would become a more efficient means of balance-of-payments adjustment than it is under the existing suboptimal configuration. It also follows that if we take the existing political boundaries as given—clearly the only realistic and relevant assumption—flexibility does not eliminate all interregional maladjustments caused by the shifts in international demand. However, no advocate of flexible rates would make such a claim. On the other hand, it does not follow from the theory of optimum currency areas that with the existing suboptimal pattern of political boundaries, flexible rates do not at least reduce, still less that they increase, the pains and losses of balance-of-payments adjustment, as compared with a regime of fixed exchanges. The reason is that monetary factors are always involved, even in cases, rare in my opinion, where the primary causal factor is a shift in international demand. Moreover, it does not follow that a real optimum always exists in the sense that it is always possible to construct currency areas in such a way that exchange rate changes would permit restoration of equilibrium without any reallocation of resources which in all probability cause at least transitional unemployment and without any changes in price levels. Clearly, in the strict sense such an optimum practically never exists, quite apart from practical-political limitations on redrawing currency areas. The reason is that some shifts in international demand are probably always present even if the primary and dominating factor is purely monetary.

---

[3] Another method which is favored by some experts would be to adjust the parity automatically at short intervals to the average of, say, the last six months.

There surely remain technical questions of details, some of which may be difficult to foresee before a system of flexible rates involving numerous countries has been put into operation. This is especially true of limited flexibility which has not yet been tried out and raises problems not involved by unlimited flexibility. Let me mention only one which is easy to foresee and would have to be dealt with before the system is introduced: It is the treatment of existing official reserves, especially dollar balances. A gold or exchange guarantee, suitably circumscribed to prevent wholesale conversion of private funds via official agencies, would probably be a satisfactory solution of that problem.

Henry C. Wallich: As a defender of fixed exchange rates, rather than expose my vulnerability on the general issue, I would like to concentrate on exchange rate policy for the United States.

U. S. Interest in Exchange Rate Stability. This country, in my interpretation, has always had a special interest in stable rates. During the 1930's, when much of the experience was gathered that crystalized in the Bretton Woods Agreements, an important concern was to protect American exports against foreign beggar-my-neighbor policies. Competitive depreciation of currencies, once the U.S. had completed its own, was one of the devices the U.S. learned to fear. As a country with a strong tendency toward payments surplus, the U.S. was in a poor position to retaliate either by controls or by devaluation. This was reflected in the determination of the American negotiators of the Bretton Woods Agreements to limit sharply the use that others could make of these instruments.

Shift from dollar shortage to dollar glut has not changed the basic case. The United States continues to be severely limited in any attempt at exchange rate mobility. Now it is principally the key currency role of the dollar that brings this about. The United States determines the price of gold. Other countries determine the exchange rate of the dollar, by pegging to it at a rate of their choice. Given that other countries have maneuverability in their exchange rate while we have not, it is clearly in our interest to try to limit freedom of exchange rate movement.

Exchange Rates and the U.S. Adjustment Mechanism. It could be argued that exchange rate movements have much to commend them as a means of adjusting the American balance of payments, in view of the obvious disadvantages of relying on income effects for that purpose. Owing primarily to the nearly closed character, on current account, of the American economy, and secondarily to our low saving and relatively low taxpaying propensities, the United States has a higher multiplier than most industrial and even developing countries. Hence, the full effect upon imports, from a given change in income, tends to be approached more rapidly abroad than in the United States. Furthermore, the low American propensity to import would require a much larger change in income to remove a given payments imbalance than it would abroad. Slow work and high cost both argue against the income mechanism as a means of payments adjustment in this country.[1]

---

[4] See especially Mundell's pioneering paper, *A.E.R.*, Sept., 1961.

The low import ratio would seem to argue that exchange rate variations should work well in adjusting the balance of payments. Demand and supply elasticities for the small volume of internationally traded goods should be very high. The range of goods capable of international substitution should be great. But once more the time factor intervenes. Long-run elasticities no doubt are high. Slight exchange rate movements should produce substantial adjustment. But this cannot be expected in the short run, thanks to the ignorance of foreign markets and foreign products natural to a nearly self-sufficient country.

Use of exchange devaluation under these conditions would require sharp overshooting to remove a given deficit quickly. Later, there would be trouble from a surplus. Whichever way one looks at it, the old saying remains true that nothing will make the American balance of payments adjust quickly.

This applies also to capital movements. It is probably an error to think that American investment abroad is decisively favored by a high exchange value of the dollar. This would be true only if American corporations paying for brick and mortar in Europe were strictly limited in their resources. If the foreign outlay is small relative to total resources, the exchange rate should not affect the investment. Whether the corporation spends $20 million to acquire a 100 million French francs property that yields 20 million French francs, or whether at a lower exchange value of the dollar, it spends $30 million, the yield is always 20 percent. By the same token, U.S. devaluation would not make American stocks more attractive to European investors, except as the risk of subsequent devaluation were reduced and as a post-devaluation boom might promise higher stock prices.

I have heard the view that this relative insensitivity of capital movements to exchange rate variations (speculative movements are something else) is precisely a reason for favoring exchange rate variability. The classical position on the relative priority of capital and current account is that capital should move freely in search of the highest yield, and that the current account should adjust. If the capital account is relatively insensitive to exchange rate movements, one way of making the current account adjust without interfering with the capital account would seem to be exchange rate variability.

I have no particular quarrel with autonomous capital movements. But I would point out that in the domestic economy, the principal vehicle of regional balance-of-payments adjustment is flows of portfolio capital. Regional imbalances in the United States are adjusted largely by flows of securities, bank-held and others. What the world needs is a supply of internationally acceptable securities to serve as secondary liquidity reserves. These would move in response to changing interest differentials, more than to exchange rate movements (except for speculative flows).

Finally, I would like to question the view that we are already practicing partial devaluation in the form of the Interest Equalization Tax, and the tying of aid and of the Defense Department dollar. I find the analogy labored. Ought we not then to regard each tariff rate as a partial devaluation? More plausibly, the Interest Equalization Tax is a tariff on the importation of securities; tying of aid and defense procurement is a vulgar export subsidy.

Floating Rates. So far I have discussed exchange rate variation without particular reference to its mechanics. I now turn specifically to floating rates.

For the United States and the rest of the world, floating rates would present an asymmetrical situation. The United States, as I said before, cannot easily determine its exchange rate. It can determine, instead, the volume of dollars the rest of the world would have to hold it if wants to peg to the dollar. There are some advantages to the United States in this situation. But basically, I fear, the asymmetry has negative effects. It is uncomfortable for the United States to be deprived of control over its rate. It is uncomfortable likewise for foreign countries to have to choose between the consequences of unlimited dollar accumulation and being undersold by the U.S. in their foreign and home markets.

---

[1] Imbalances between the U.S. and the rest of the world are reduced, of course, at the same speed.

Unless American policy is able to inspire great confidence, therefore, I would expect foreign countries to react to floating rates by protecting themselves. They might split the dollar rate, pegging the trade dollar and letting the finance dollar find its own level, as the Swiss did at the end of World War II. Or they might introduce direct trade controls to protect at least their home markets against cheap American competition.

I know, of course, that under a truly flexible system countries are not supposed to influence their rates. The rest of the world ought not to peg to the dollar. The United States ought not to do anything to push the dollar down. In fact, I would expect floating rates to cause serious conflicts among countries, each trying to influence its rate. The proposition that there should be no intervention, except possibly at the edges of a band, appears to me implausible. If official intervention in the exchange market were nevertheless effectively outlawed, intervention through interest rate policy, forward rate policy, and through semiofficial agencies would remain possible, in some cases inevitable.

It has been said that flexible rates pose a threat principally to the international flow of goods and capital, because of uncertainty. I suspect they might turn out to threaten also domestic investment in countries in danger of appreciation, unless these countries offer protection to the domestic investor. To the country whose currency is falling, the situation is pleasant. To the country that, for reasons that may have nothing to do with the state of its current account, finds its rate floating upwards, the situation may be intolerable.

The threat I am describing can possibly be met by limiting flexibility within a band—if the band holds. But if countries inflate at differential rates, one edge of the band will simply become the new parity or peg. And if there is not enough financial discipline to hold the original parity, what is to sanctify the surrogate? The United States, moreover, so long as the dollar is the reserve currency, cannot even control the band of the dollar.

A moving band, with a moving parity based upon an average of past rates, has been suggested to meet this difficulty. The device would not limit the extent of rare movements, but their speed. I would suspect that it would add to that speed. A country faced with the alternative risks of not being able either to appreciate or to depreciate as much as it would like, probably will consider the latter the greater danger. It would then be motivated to keep its rate on the low side, to bring down the moving average, thus engaging in precautionary depreciation.

PETER B. KENEN: Discussions of the choice among exchange rate systems should address themselves to several questions, and those questions should be ordered in logical sequence. I shall pose three such questions and shall offer three tentative answers, each of which should provoke a new set of questions. My three questions, put quite simply, are these:

First, is there any way to know when exchange rates should be changed and when they should be held against a tendency to change?

Second, what is the cost of failing to change rates when a change is needed and the cost of changing them when no change is needed?

Third, who is more likely to detect the need for change: central bankers or private traders?

Too much of the debate on flexible exchange rates has turned on the third question and focused still more narrowly on the role of speculation. Much has been written, theory and history, on the first and second questions, but we have not made sufficient reference to those same two questions in debates about exchange rate systems.

Our knowledge of the circumstances in which change is needed can best be summarized by turning on its end Johnson's important distinction between expenditure-changing policies and expenditure-switching policies. Ignoring, for the moment, capital flows (which call for expenditure switching to alter the trade balance; then for expenditure changing to restore internal balance), one can pair these policies to the disturbances afflicting the exchange rates.

Suppose that all countries are at full employment (with stable prices) and exchange markets are in equilibrium without official intervention. If one country inflated, it would suffer rising prices and, with static-stable foreign exchange markets, its currency would depreciate. That depreciation would serve one useful function: it would bottle up the national inflation, sparing other countries any need to change their domestic policies. But it would not solve the problems of the country with inflation. Instead, the bottling up of its own inflation would cause its citizens to feel the full effect of its mistake, and the rise in import prices due to depreciation might evoke additional cost-generated price inflation.

Some would argue, righteously, that this is the proper outcome. Exchange rate flexibility confines a country's errors to its own boundaries and punishes the voters for their government's mistakes. But we would all agree that there is a better way to deal with such situations: to use domestic policies to regulate demand. In brief, an expenditure-changing disturbance can best be opposed by an expenditure-changing policy.

I will not prolong my analysis by offering a second, symmetrical example, to show that expenditure-switching disturbances call for expenditure-switching policies, but will simply state the general result: If governments were able to control domestic demand (to maintain full employment and also to keep pace with price drift elsewhere in the world), they would not have to use changes in exchange rates unless they did encounter nonreversible switches in patterns of expenditure (including those attending economic growth) or comparable "structural" changes in nonmarket transfers and capital flows.

Notice, moreover, that changes in exchange rates are not the only optimal responses to such disturbances. Factor mobility is, as Mundell shows, a satisfactory substitute for change in the terms of trade of the type implied by changes in exchange rates. And in very small, open economies, one might even opt for selective measures, for reasons adduced by McKinnon and others. We are, by now, familiar with the theory of optimum currency areas (and with the need to adapt that theory to the real world problem of grouping existing countries into the appropriate fixed exchange rate zones).

One could say much more about my first question, but our Chairman is well known for watching the clock, and I have now to turn to my second question: to ask about the cost of changing exchange rates when they should be held and the cost of not changing them when they should be altered.

My earlier example gives a partial answer. If exchange rates change because of an inflation, the erring country suffers more and other countries suffer less (or not at all). This is not unjust, but neither is it neighborly. If a single country were the sole significant villain in these matters, other countries should not have to share the costs of its folly. But all countries make mistakes from time to time, and it may be most neighborly to share each others' errors, thereby to lighten the burden of the error each country has to bear when it has made the mistake. Furthermore, I am impressed with Triffin's frequent point—that the cost inflation attending depreciation may increase the average rate of price inflation all over the world, because a depreciation may push home prices up more than the corresponding appreciation of other currencies pushes foreign prices down.

Yet one must entertain a countervailing argument. Changes in exchange rates, even when second best, are better than the policies usually pursued when a country fails to regulate demand. The tendency to use controls—to manipulate implicit multiple exchange rates—may be more costly than a second-best resort to an overt change in the explicit rate, because selective measures distort resource allocation. This point needs no long elaboration at a time when the United States ties foreign aid, puts a hidden tariff on government procurement, taxes acquisitions of foreign securities, and directly regulates other flows of capital.

The same point, however, has also to be turned around. Governments may opt for selective measures when, in fact, they do confront disturbances which call for exchange rate changes. One might argue, indeed, that American experience through 1964 resulted from a series of "structural" disturbances, in market and government transactions alike, that called for substantial expenditure switching, and that the policies actually adopted comprised an inappropriate combination of expenditure reduction (or too cautious stimulation) and inefficient expenditure switching.

I come, now, to the third of my three questions, relating directly to flexible exchange rates. Here, I can add little to what is now known, save to reformulate the principal issue. One must ask who is most likely to use his cash reserves to keep exchange rates stable when they should not change and who will facilitate exchange rate changes when they are appropriate. This, indeed, is the question posed by the classic controversy about private speculation. Under fully flexible exchange rates, speculators take the place of the central banks; they commit their cash reserves to hold rates still, to damp fluctuations, or to hasten change, according to their judgments concerning future rates. One might argue that their judgments are liable to distortion because they are influenced by other speculators' views, and that average opinion about average opinion does not always generate ideal prices. One might argue, contrarily, that Canada's experience shows that the speculators are not quite so silly and do, indeed, stabilize a viable rate. But Wonnacott's analysis convinces me that this was a special case, strongly affected by an impressive history of exchange stability and by the so-called "parity psychology."

I would, myself, predict that flexible rates would move too often—when change is suboptimal as well as optimal—even as pegged rates move too infrequently.[1] But I do not know which bias is worse.

To close, I want to raise one important question that has been too much argued and too little analyzed. Opponents of full flexibility, especially bankers and businessmen, argue that it would constrict foreign trade and investment because it would generate undue uncertainty concerning future rates. Proponents reply that forward markets would soon develop, even in the very long maturities required for trade in capital goods. But neither side has raised the basic question: which system would generate the greater fear of changes in exchange rates not too far ahead?[2] It is this fear of change, not the recorded facts, that inhibits the development of long-term forward markets and, to this extent, inhibits foreign trade. Comparing the experience of Britain and Canada, one might perhaps indict the pegged-rate system. But recent experience with changes in parities—in 1961 and 1967—should modify our preconceptions, and the growing evidence of high price elasticities suggests that either system can be made to work quite well with moderate exchange rate changes.

The issue, in the end, is horribly subjective. Will it be easier to persuade the central banks that speculators are not gnomes and that they can be trusted to manage the exchange rates, or instead to convince them that their own self-esteem does not depend upon the preservation of a fixed parity, but rather on their skill in choosing a new rate?

Milton Friedman: Those of us who have favored a regime of floating exchange rates are repeatedly met with one variant or another of "it takes two to tango." We are told that, as a result, a system of floating rates will require as much international cooperation as a system of fixed rates, that those of us who are desirous of limiting rather than extending the power of international organizations of central bankers are visionary in supposing that the adoption of a system of floating exchange rates will do so. Finally, we are told that in advising the U.S. to float the dollar, we are "unrealistic" because other countries will not permit us to do so.

---

[1] This is, of course, quite different from the charge that pegged rates move too much when, in fact, they move at all. But this charge may be answered by the two most recent changes in key exchange rates: the deutschemark-guilder revaluation of 1961 and the sterling devaluation of 1967. Its limited validity derives from the experience of the 1930's and 1949.

[2] Note that the much discussed abruptness of pegged-rate changes is not the issue here. Gradual change in a flexible rate over four years and one sudden large change at the end of four years pose similar problems for a trader or investor who confronts a four-year contract now.

The discussion of these issues has been confused on both sides—and I plead guilty to contributing to this confusion—by failure to keep sharply separate the options that are available to a single country and those that are available to the international community. Clearly, for domestic and international political reasons, a country may not wish to exercise its options. But that is a very different thing from what it can do on its own if it wishes.

The critics are right that the U.S. cannot on its own float the dollar in the fullest sense of that term. Hence, I no longer describe my policy recommendation for the U.S. in those terms. What the U.S. alone can do—and what I continue to believe it should do—is set the dollar free by stopping pegging the dollar. It can leave it up to other countries whether the dollar floats or whether they link their currencies to the dollar.

Consider the following steps by the U.S.: (1) cessation of gold purchases or sales at $35 an ounce or any other price; (2) elimination of all exchange controls and restrictions—repeal of the interest equalization tax, elimination of controls over bank lending and private investment, of "voluntary" and involuntary quotas, and so on; (3) an announcement that neither the Treasury nor the Federal Reserve will intervene in markets for foreign exchange to affect its price, that the only foreign exchange operations engaged in by the U.S. government will be those that arise in the course of other activities (e.g., purchases of foreign exchange to pay for goods and services acquired abroad).

Each of these steps is within the unilateral control of the U.S. No other country can by its action prevent us from taking them—just as none could prevent us from going farther and removing all tariffs.

Whether the consequence of these actions would be a system of freely fluctuating exchange rates depends on what other countries would do. Other countries might peg their currencies to the dollar, thus creating a dollar standard or a dollar bloc characterized by fixed exchange rates. How could they do so? Only by (1) adding to dollar holdings when they are in surplus and providing their currency in exchange, or drawing down dollar reserves when they are in deficit; or (2) using import, export, and exchange controls to prevent unduly large surpluses or deficits at the fixed exchange rate; or (3) adjusting their internal monetary policy to harmonize with U.S. policy.

So far as the U.S. is concerned, I do not see why we should object to any of these steps. The first would involve an increase in dollar loans to us at zero or low interest rates when the U.S. is in deficit, which we certainly have no reason to object to, and a repayment of these loans when we are in surplus. On balance, it could be expected to mean a larger average volume of dollar liabilities (i.e., of low-interest rate loans) because of the greater need for dollar reserves it would generate.

The second involves measures of a kind we disapprove of. However, in the context of an adjustment to our abolition of restrictions, it means at worst that foreign countries substitute their restrictions for ours. It does not mean any net increase in restrictions, because there is no reason why the foreign countries should seek to reduce their surplus to a lower level than they are now willing to accept.

The third is clearly not objectionable to us. It simply enlarges the dollar area.

So far as foreign countries are concerned, circumstances differ very much from country to country. For many countries, it would make a good deal of sense to link their currencies to the dollar. These are countries that have a large foreign trade sector, are closely linked economically to the U.S., and are not likely on their own to have a more effective internal monetary policy than the U.S. For larger countries, such as Britain, Germany, and France, I believe they would be better advised in their own interest to follow the same policy as I have suggested for the U.S. This would mean freely floating rates between their currencies and the U.S.

What would in fact happen need not of

course correspond to what is desirable, and, even less, to what I regard as desirable. This is unexplored territory in which prediction is treacherous. My own conjecture is that many countries would peg their currencies to the dollar but that the Common Market countries would not be willing to do so and would at least try to set up a kind of gold bloc whose currencies would be linked at fixed rates to one another but fluctuate as a group vis-à-vis the dollar. The pound might form a third main currency bloc or might be pegged to the dollar.

I am very uncertain about the precise form these developments would take. But certain things do seem clear. First, if the U.S. took the steps suggested and if it followed a reasonably stable and noninflationary internal monetary policy, the dollar would continue to be the major vehicle currency employed in international trade—indeed would grow in importance. Second, New York would resume its growth as the center of the capital market of the world. The interest equalization tax and foreign exchange controls (assisted by ceilings on interest rates banks may pay on deposits) have interrupted this development. They have artificially stimulated the Euro-dollar market and kept London alive as a major capital market. Set the dollar free, and make it clear that it will continue to be free, and the Euro-dollar market will shrivel and New York will become in the final decades of the twentieth century what London was in the final decades of the nineteenth century.

It is one of the ironies of history that the New York banks have themselves inhibited this development by their stubborn insistence that the U.S. peg the price of gold and the dollar.

Fritz Machlup: Most discussions of the comparative advantage of fixed and flexible exchange rates suffer from semantic confusion. This is almost unavoidable, because so-called "fixed rates" are frequently altered and thus in effect not only quite flexible but often also very unstable in the long run; flexible rates, on the other hand, may in fact be relatively stable in the long run.

For example, in the twenty years from 1947 to 1967 fixed par values were changed twice for the German mark, twice for the pound sterling, and five times for the French franc; and the spreads between the highest and the lowest values were 21 percent for the mark, 40 percent for sterling, and 76 percent for the franc (calculated on the basis of the highest parities, which is the more charitable technique of calculation). In contradistinction, the Canadian dollar, which had a flexible rate for twelve of the same twenty years, had a maximum spread of only 18 percent of its highest rate. Thus, we must carefully avoid the impression that fixed rates are stable and that flexible rates are unstable. It was, I suppose, only inadvertence when one of our four panelists today used the illegitimate dichotomy "fixed versus unstable." Fixed and stable are neither logically nor empirically correlated.

We must make a conscious effort to fix and stabilize the meanings of the terms which have been used all too flexibly in the discussion of our subject over the years. Instead of fixed rates we should perhaps say "fixed par values" and recognize that they usually are fixed only until the balance of payments is in such a fix that the par values have to be abandoned, usually to be refixed at another level. In a recent international discussion among some very high-powered experts, all protagonists of fixed par values conceded that this system, if it is to work properly, requires adjustment of the official pegs for major currencies as often as every five or six years. We should also remember that in the 1940's some of the participants in and several commentators on the Bretton Woods Agreement called the system one of exchange-rate flexibility. This was justified by the fact that the covenant provided for changes in par value, small changes at the discretion of the member country and larger changes with approval of the Fund in situations regarded as fundamental disequilibrium.

In view of this terminological history and particularly in view of the frequency and amplitude of changes in official par values, I would prefer to discard both words "fixed" as well as "flexible." I have learned, however, that the excommunication of terms is not accepted, even if communication is seriously encumbered by their continued use. All I can hope to do is to suggest qualifications and distinctions that would reduce the danger of confusion.

In common parlance, the system of fixed rates is one in which countries peg the exchange rate of their currency close to (that is, within 1 or 2 percent of) a formally announced par value; the term "fixed" is applied regardless of whether the official peg is adjusted every fifteen years, every five years, or every year. The opposite of "fixed rates" allows of so many variations that any definition that encompasses all varieties would be too clumsy to serve a good purpose. It is preferable to provide classifications that point up the major differences. Analysts not making these distinctions may be trapped into offering broad generalizations that are not valid for all types of flexible-rate systems.

The following classification is based on the scope of official management and on the limits set to the variability of the exchange rate:

	Unlimited Variability	*Limited Variability* *around Parity* ("wider band")	*of Parity* ("crawling peg")
Unmanaged	Freely flexible (no interventions)	Interventions only at the edges of the band	
Managed	(a) by discretion of national authorities (b) by international agreement (c) by discretion of international authority	Interventions also within the band	(a) by discretion of national authorities (b) by international cooperation (c) by fixed formula

The most fundamental distinction is between freely flexible or floating rates, on the one hand, and rates with limited and managed flexibility, on the other hand. Managed flexibility exists when monetary authorities intervene in the foreign-exchange market by selling foreign currency to lower its price or prevent it from rising or by buying foreign currency to raise its price or prevent it from falling. The distinction is crucial for several reasons; one of these is that the sale of foreign currency by the monetary authorities takes domestic money out of circulation or cancels it, and that the purchase of foreign currency by the authorities increases the domestic money supply. Thus, for example, if a country devalues its currency, the central bank, as it purchases foreign exchange and pays a higher price for it, supplies additional domestic money. If, on the other hand, the currency of a country depreciates in the foreign-exchange market with no official purchases or sales of foreign money, the higher price for the foreign money is paid by private buyers, who cannot increase the supply of domestic money, or at least not the money base by which the money supply is limited. Unless commercial banks increase their long position in foreign balances, the fall of a floating rate of domestic currency in the free exchange market will not increase, and may reduce, domestic liquidity. Nonawareness of this difference between a flexible rate with official interventions in the foreign-exchange market and a "floating" rate marks insufficient understanding of the underlying principles.

If we sometimes hear official experts express the fear that a system of floating exchange rates will lead to competitive devaluations, we may pardon them for this terminological slip. They probably do not mean floating rates, but mean to refer to managed rates, managed in the spirit of the 1930's, without international cooperation, when expansion of exports was widely regarded as the only practical way to increase employment, and when countries accordingly attempted to export their unemployment to one another. Governments, nowadays, no longer try to combat unemployment by deliberate devaluation; the fears expressed regarding such policies reflect poor political judgment. Competitive devaluation designed to help the country compete for a larger share in a given volume of world employment is a thing of the past. There may still be a danger of serial devaluations if countries have to compete for desired shares in a given volume of world reserves, but deliberate creation of new reserve assets is supposed to take care of official liquidity preferences. In any case, the terminological confusion between a system of competitively adjustable pegs and a system of floating exchange rates should be avoided.

A more embarrassing mistake can be found in the frequently repeated remark that floating rates are likely to be sinking rates. It should be obvious that no currency can depreciate relative to other currencies without these other currencies appreciating relative to the former. If some rates go down, others must go up (unless all movements are measured against some "standard," a currency or commodity chosen as a numeraire). Commentators who expect that their currency will sink when it is left to float without official support evidently judge it to be currently overvalued or they believe that the authorities of their country are more susceptible to inflationist pressures than those of other countries.

Speaking of the propensity to inflate, it is probably true that devaluation—that is, a deliberate action of the authorities, to pay a higher price for foreign currency—is more likely to raise the voices and increase the pressures of the advocates of monetary expansion than depreciation of the currency, which is not a deliberate action of the authorities but an anonymous reflection of supply and demand in the free market. I side with Haberler (and with Kamitz and Schmitz, two Austrian central bankers) in the belief that a decline of the exchange rate of one's currency may be a more alarming signal for the authorities than a decline in the official reserves and, hence, may induce more rather than less discipline in the exercise of monetary policy. I realize that some central bankers may have to learn a thing or two, but I do not share the view that they are not teachable.

I realize that the prejudices of the financial community against freely flexible rates are too strong to be overcome in the near future. Thus, limited flexibility is all that we can expect to get adopted. We must distinguish limited variations around a fixed peg, limited variations of the peg, and the combination of these two; that is, limited variations around a slowly adjustable or slowly self-adjusting peg. The term "sliding parity" (proposed by James Meade) may create in the minds of some the false connotation of a downward slide. The term "crawling peg" (proposed by John Williamson) leaves it uncertain whether the crawl is dictated by the authorities or left to the automatic operation of a fixed formula. The term "self-adjusting peg" (proposed by Hendrik Houthakker) is supposed to denote the variability limited by a fixed formula, in particular by the average of the exchange rates of the past months (as in the proposals of Carter Murphy and John Black).

Our four panelists today may be ranked according to their inclinations toward flexibility or rigidity of exchange rates. Friedman prefers freely flexible rates, without any official interventions. Haberler recommends limited flexibility in the form of a wider band around a parity that is subject to slow, continuous adjustments. He is not specific on whether he would want authorities to intervene only when the limits of the band are reached or also within the band. Kenen recognizes the need for variability of exchange rates but seems to favor discreet changes over continuous ones. He thinks that the social costs of making adjustments that later prove to be unnecessary exceed the social costs of postponing needed adjustments. Wallich is the least flexible of the group, chiefly because he fears that in practice the wider band will always mean clinging to one of the two edges of the band and that the crawling peg will usually mean a crawl toward competitive depreciation.

This looks like a wide spectrum of professional opinion. I suspect, however, that if we persuaded the four men to agree on some compromise, Friedman might settle for Haberler's position and Wallich for Kenen's. The difference then would reduce to the question whether abrupt changes or continuous changes in exchange rates are more wholesome. To make things quite concrete, I ask two questions: Would it have been better if the pound sterling had declined by no more than 2 percent a year over six years or was it better to have it propped up at a high level over several years, with a series of emergencies and rescue missions, to be finally reduced in one fell swoop by 14 percent? Is it wiser to keep the German mark undervalued at its fixed parity with periodic rumors that it will over the next weekend be revalued by 7 1/2 or 10 percent or would it have been preferable to let it go up by 1 or 2 percent a year, as the German Council of Economic Advisers has consistently recommended for several years? To your immoderate moderator the answers are clear.

If the four panelists allowed me to moderate their conflict, I would strongly argue that a wider band of exchange rates around a slowly crawling peg is for several countries—not for all, of course—a much more workable system than the present with the narrow band around a temporarily rigid peg that is abruptly adjusted when massive speculation makes its position untenable. I argue this chiefly on two grounds: (1) Speculative capital movements are much more likely to occur when one can make a gain of 10 percent over a weekend than when one can make at most 2 percent over a year. (2) The real adjustment process is set in motion by a small exchange-rate variation slowly, almost imperceptibly, yet years before it can start if the adjustment of the parity is delayed. In my opinion flexibility provided by a wider band around a slowly self-adjusting peg is preferable both on capital account and on trade account to the spasms associated with a system of rigidity with abrupt and drastic corrections. The system of fixed par values is in effect a system of "jumping pegs."

ECONOMIC HISTORY

THE COST TO AMERICA OF BRITISH IMPERIAL POLICY*

By PETER D. MCCLELLAND
Harvard University

I

The American Revolution—what Adam Smith preferred to speak of as "the late disturbances in the colonies"—has long remained a central puzzle of American historiography. In the maelstrom of events that led to Lexington and Concord the role of economic factors is still a subject of dispute. That Britain repeatedly sought to regulate production and trade within the colonies can easily be documented, as can the resentment that such regulations evoked throughout her American possessions. What remains obscure is the motivation for colonial indignation. Was British interference resented primarily because it jeopardized colonial freedoms or because it jeopardized colonial pocketbooks?

Essential to resolving such a dispute is some indication of the actual costs involved. Historians have repeatedly assailed the problem by listing all relevant imperial policies, noting whether the tendencies of each was to help or to hinder colonial development, and then—given that list of tendencies—reaching some vague conclusion on the net impact of all combined. Not surprisingly those conclusions have varied widely. Curtis Nettels, for example, asserts that imperial restraints were "restrictive, injurious, negative"[1] while Oliver Dickerson claims that they were not.[2] The solution to such a dispute is obvious enough. What must be assessed is not just the directional impact of each policy, but the dollars and cents involved.

Only one aspect of British imperial policy has been subjected to cost-benefit analysis of this sort. In 1939—and more recently in 1965—attempts were made to measure the costs to the colonies of trade route distortions attributable to the Navigation Acts.[3] Here, too, the dispute continues unresolved. The estimates of Lawrence Harper imply that British imperial policy placed "a heavy burden upon the colonies"[4] while those of Robert Thomas suggest the absence of any "significant hardship."[5]

This article has a twofold purpose. The first is to show that the conceptual framework of both Thomas and Harper is so defective that their final results are virtually meaningless. The second is to offer an alternative upper-bound measure of the cost to America of the British Acts of Trade.

Before the cost of British domination can be appraised, the alternative to such domination must be clarified. If that alternative is American independence, the problem becomes to decide when the thirteen colonies could first have resisted the territorial ambitions of all European powers. Had British rule been overthrown early in the eighteenth century, the fleur-de-lis might well have supplanted the stars and stripes along the eastern seaboard of North America. Only with Britain's victory in the Seven Years' War did France cease to be a power of importance in the New World. Counterfactual analysis that would contrast British rule with the hypothetical alternative of American independence would therefore seem best confined to the thirteen-year period between 1763 and 1775.[6]

---

* The writer is indebted to James Shepherd, of Purdue University, for supplying the trade data estimates indispensable to the computations of Appendix A.

[1] Curtis P. Nettels, "British Mercantilism and the Economic Development of the Thirteen Colonies," *J. of Econ. Hist.*, Spring, 1952, p. 114.

[2] Oliver M. Dickerson, *The Navigation Acts and the American Revolution* (New York: A. S. Barnes & Co., 1963), p. 52.

[3] Lawrence A. Harper, "The Effect of the Navigation Acts on the Thirteen Colonies," *The Era of the American Revolution*, ed. Richard B. Morris (Columbia Univ. Press, 1939), pp. 3–39; and Robert Paul Thomas, "A Quantitative Approach to the Study of the Effects of British Imperial Policy Upon Colonial Welfare: Some Preliminary Findings," *J. of Econ. Hist.*, Dec., 1965, pp. 615–38.

[4] Lawrence A. Harper, "Mercantilism and the American Revolution," *Canadian Hist. Rev.*, Mar., 1942, p. 3.

[5] Thomas, *op. cit.*, p. 616.

II

The Navigation Acts by the time of their repeal embraced over 340 different pieces of legislation. Imbedded in that morass of shifting rules and exceptions can be found two major provisions that directly affected mercantile activity in British North America. All trade between England and her colonies was restricted to vessels built and owned within the Empire, with crews predominantly of colonial or British origin. In addition, much of the trade between colonial possessions and foreign countries was artificially routed through English ports, thereby strengthening the entrepôt position of the mother country. To assess the cost to America of this second provision has been the major (although by no means the only) task of Harper and Thomas.

What is to be measured is portrayed in Figure 1. Consider the case of tobacco: the dominant American export subject to enumeration. Let:

S_{fob} = the supply curve of colonial tobacco, f.o.b. Maryland
D_{cif} = the European demand curve for Maryland tobacco, c.i.f. Amsterdam
T = the total cost per pound of moving tobacco from Maryland to Amsterdam by way of Britain
D_{fob} = the European demand curve for Maryland tobacco, minus the cost of moving tobacco from America to Amsterdam
$= D_{cif} - T$

The intersection of D_{fob} with the colonial supply curve S_{fob} determines the quantity exported (Q_1) and the price f.o.b. Maryland (C_1). The Amsterdam price (A_1) is simultaneously determined by ($C_1 + T$). Let:

ΔT = the incremental costs incurred by shipping tobacco indirectly via an English port rather than directly from America to Amsterdam
D'_{fob} = the European demand curve f.o.b. Maryland that would have prevailed without these incremental costs
$= D_{cif} - [T - \Delta T] = D_{fob} + \Delta T$

The intersection of D'_{fob} with S_{fob} demonstrates that without trade route distortions caused by the Navigation Acts, the colonists would have sold more tobacco (Q_2) at a higher price f.o.b. Maryland (C_2), even though the prices in Amsterdam would have fallen to $A_2 (= C_2 + [T - \Delta T])$. The associated rise in American incomes is indicated by the area C_2C_1KL.

Similar analysis can be applied to all colonial exports and imports artificially routed through Great Britain by the Navigation Acts. (In the case of imports, the European supply curve f.o.b. can be inflated by the costs of transatlantic shipment to derive a supply curve c.i.f. America. The intersection of this inflated supply curve with the colonial c.i.f. demand curve will then indicate prices in the colonies and quantities imported.) In each case, the burden of imperial policy

---

[6] See Thomas, *op. cit.*, p. 617. As Douglass North has pointed out, a major defect in Harper's analysis is his failure to specify a consistent hypothetical alternative. Douglass C. North, *Growth and Welfare in the American Past* (Prentice-Hall, 1966), p. 47.

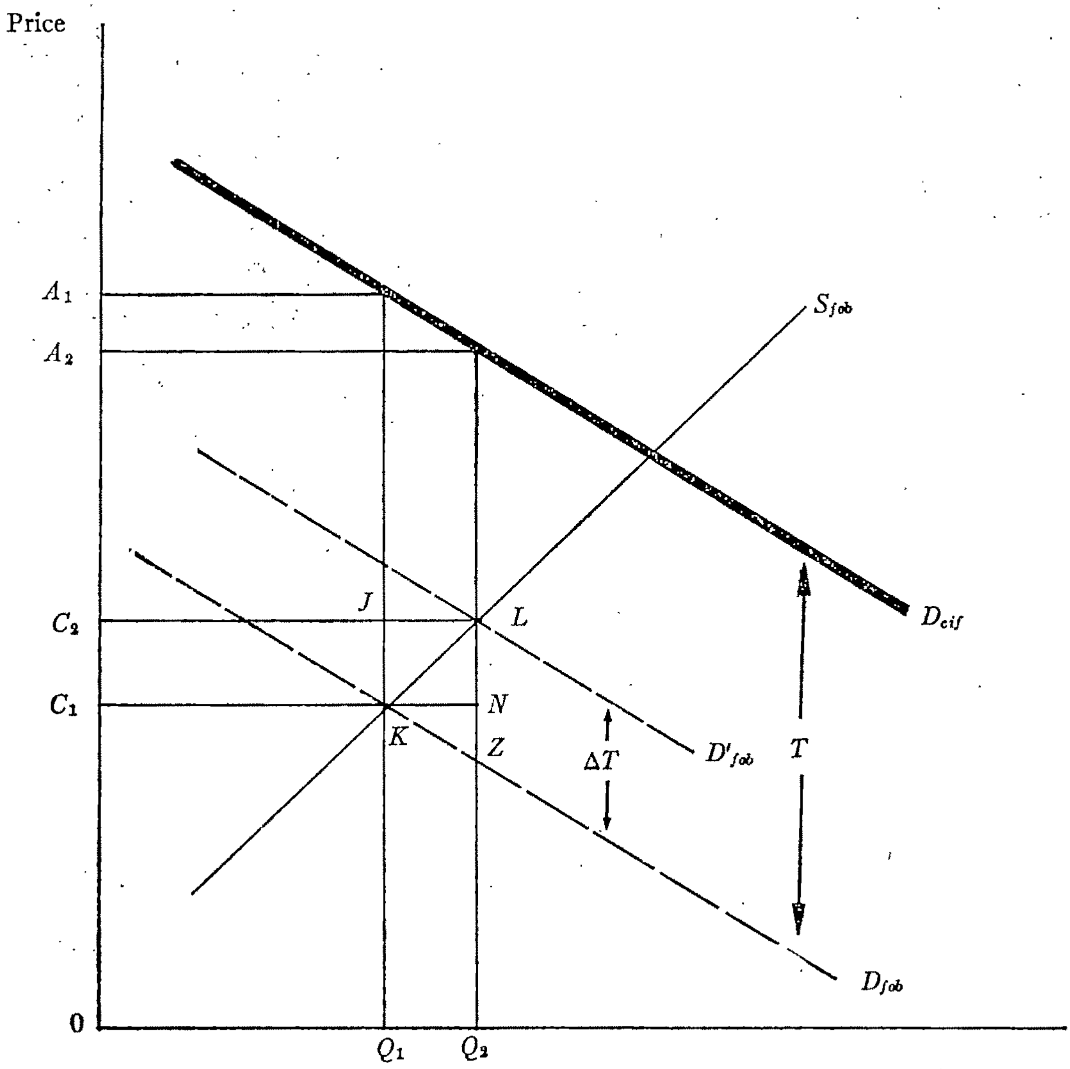


FIGURE 1

$$= C_2C_1KL$$
$$= (C_2 - C_1)Q_2 - KNL$$
$$= (C_2 - C_1)Q_1 + KJL$$

can be approximated by either $(C_2-C_1)Q_2$ or $(C_2-C_1)Q_1$. Which approximation is more suitable depends on the case that one wishes to make. If the objective is to stress the debilitating effects of British interference, the lower-bound measure of $(C_2-C_1)Q_1$ should be used. A large result then guarantees that the true costs were larger still. The minimum information for such a study includes not only data on the prices of colonial exports and imports and quantities shipped between 1763 and 1775, but also estimates of what prices would have been in the absence of British regulations (C_2). Such is the challenge faced by Lawrence Harper. For Robert Thomas the task is even more complicated. His goal is to minimize the importance of British interference. He must therefore employ the upper-bound measure of $(C_2-C_1)Q_2$, a measure that requires knowledge about both prices (C_2) and quantities (Q_2) in a world that never was: the world of American independence, 1763 to 1775.[7]

One proxy for this world is the world of actual American independence after 1783. Both Harper and Thomas use the latter to estimate what colonial trade would have been in the former. Three obstacles must be overcome—all of formidable dimensions. Prices from the Federalist era cannot be compared with those of colonial days without some adjustment for the inflation that occurred in the interim. Even if this is successfully accomplished, the end result may have little meaning. Prices in post-revolutionary American were determined by the intersection of supply and demand schedules of the sort portrayed in Figure 1. To use these numbers to approximate what prices would have been had America been independent two or three decades earlier is to assume that neither supply nor demand schedules shifted in the intervening period. That such rock-like stability should have pervaded the American foreign trade sector for over a quarter of a century appears to be dubious in the extreme. But assume that it did. How could one assess the importance of the final measure of C_2C_1KL? For purposes of deciding its aggregate impact on colonial economic growth, the obvious touchstone is gross national product. No detailed estimates of GNP have been made for colonial America, nor are they likely to be. The necessary data simply do not exist. Without such a frame of reference, one is hard pressed to judge the significance of income lost because of trade route distortions attributable to the Navigation Acts.

Harper's analysis founders on the first two difficulties. To adjust for inflation, he corrects prices from a wide range of periods by 78 percent—Bezanson's estimate of the general price rise between 1741–45 and 1821–25.[8] So diverse are the periods from which he assembles data that his final cost estimate is valid only if supply and demand schedules remained unchanged for over a century. In the case of tobacco, for example, "surprising unanimity of result" is achieved in three independent estimates of (C_2-C_1).[9] One of these—the differential in Philadelphia prices between 1720–75 and 1784–1861 adjusted for a 78 percent change in the price level—is suspect for reasons already suggested. The other two attempt to measure not (C_2-C_1) but ΔT: the incremental costs of moving goods by way of Great Britain. As indicated in Figure 1, ΔT is an upper-bound measure of (C_2-C_1), a bias that is inconsistent with Harper's objective of a lower-bound measure of C_2C_1KL. His first estimate of ΔT is based on the 1720 testimony of two tobacco merchants[10] concerning the differential between English reexport prices and colonial prices—a differential that measures all transatlantic costs, not incremental costs (ΔT). The second estimate is based on the differential between British import and reexport prices, the notorious official values of Great Britain being corrected by reference to Moreau's work,[11] even though the same correction procedure (by Harper's own admission) yields nonsensical results in the case of rice.[12] Faced with such defects, the reader feels compelled to search for alternative cost estimates based upon less controversial procedures.[13]

---

[7] Thomas actually attempts the more complicated estimate of $(C_2-C_1)Q_1+KJL$ by employing various assumptions about the price-elasticity of supply. See, for example, Thomas, *op. cit.*, p. 624.

[8] Harper, "The Effect of the Navigations Acts," p. 33. Harper inflates C_1 rather than deflating C_2, thereby giving to the differential (C_2-C_1) an upward bias that is inconsistent with his lower-bound objective.

[9] *Ibid.*, p. 34.

[10] *Ibid.* Harper actually cites Lewis Gray (*History of Agriculture in the Southern United States to 1860*) who in turn refers to Great Britain, *Journal of the Commissioners for Trade and Plantations, November 1718–December 1722*, p. 210.

[11] César Moreau, *Chronological Records of the British Royal and Commercial Navy* (London, 1827). Moreau (p. 23) calculates a twelve-year average, 1763 to 1774, for official and real values of British exports and imports by broad geographical regions, real values "com-

The procedures of Robert Thomas are, if anything, even more controversial. Much of his analysis rests upon price comparisons for four commodities: tobacco, rice, tea, and pepper. In each case, the following adjustments are made. If:

A_1 = the Amsterdam price of Virginia tobacco before the American Revolution

A_2 = the Amsterdam price of Virginia tobacco after the Revolution

C_2^* = the Philadelphia price of tobacco after the Revolution

then

C_2^* = Thomas' estimate of what Philadelphia prices would have been in a free America, 1763–75
$= C_2^*(A_1/A_2)$

The purpose of this adjustment is to derive a prewar price that will reflect the advantages the colonists would have received "by shipping directly to northern Europe rather than indirectly through England."[14] If American price inflation had not occurred (and supply and demand schedules had remained constant except for the shift portrayed in Figure 1) those advantages are already portrayed by prices in postwar Philadelphia, with no further adjustment necessary (i.e., C_2^* is equivalent to C_2 in Figure 1). Postwar American prices must not be corrected for the reasons suggested by Thomas. They must be corrected for American price inflation that occurred between the pre- and postwar eras. The empirical issue is whether the Thomas procedures (albeit by accident) are an adequate proxy for the latter.

Suppose for example that the American prices of 1785–89 are assumed to be indicative of what prices would have been had America been independent between 1770 and 1774. According to the indices of Bezanson[15] or Warren and Pearson,[16] prices of 1785–89 should be deflated by 5 percent to 10 percent to make them comparable to the price levels of 1770–74. Similar adjustments by the Thomas procedure involve deflating American tobacco prices by 16 percent, pepper prices by 18 percent or 25 percent (depending on whether one uses the Amsterdam price ratio for black or white pepper) and either inflating American tea prices by 29 percent or 52 percent or deflating them by 38 percent (depending on whether one uses the Amsterdam price ratio for Souchon, Buoy, or Peco tea).[17] If instead of 1785–89, the period 1790–94 is contrasted with 1770–74, the associated A_1/A_2 ratios involve an even wider range of inflations and deflations.[18]

Amsterdam price changes during this period were the net product of general price inflation within Holland[19] as well as a

---

puted at the average Price Current." For details of Harper's use of these data, see Harper, *op. cit.*, p. 33, n. 86.

[12] Harper, *op. cit.*, p. 35.

[13] Other criticisms of Harper's work may be found in Douglass C. North, *Growth and Welfare in the American Past*, pp. 47–49.

[14] Thomas, *op. cit.*, p. 624.

[15] Anne Bezanson *et al.*, *Wholesale Prices in Philadelphia* (Univ. of Pennsylvania Press, 1936), p. 388 (20 commodities) or p. 390 (12 commodities).

[16] U. S. Bureau of the Census, *Historical Statistics of the United States* (Washington, D.C., 1960), p. 116.

[17] Amsterdam prices from N. W. Posthumus, *Inquiry into the History of Prices in Holland* (Leiden: E. J. Brill, 1946), Vol. I, pp. 176–80, 190–94, 203. Rice estimates are confounded by a shortage of prerevolutionary data for Amsterdam, Thomas relying on 1760 prices to effect all of his calculations. Thomas, *op. cit.*, p. 625.

[18] These results are (negative signs indicate deflation): tobacco, −5 percent; black pepper, −33 percent; white pepper, −26 percent; Buoy tea, 94 percent; Souchon tea, 64 percent; Peco tea, −31 percent.

[19] Cf. Posthumus, *op. cit.*, I, p. ci.

multitude of shifting supply and demand conditions for specific commodities. One of those shifts resulted from the more direct routing of American trade after the revolution, as portrayed in Figure 1. That this was the only shift in supply and demand schedules during the entire period under study is belied by the very diversity of the A_1/A_2 ratios. The estimates of Thomas are therefore defective on the same two counts as were those of Lawrence Harper: the questionable procedure used to correct for American price inflation; the assumed constancy in supply and demand conditions for over a quarter of a century.

III

The challenge—to repeat—is to measure the percentage of colonial gross national product that was sacrificed because of trade route distortions caused by the Navigation Acts. One obvious approximation is:

$$\frac{(x-j)(\sigma)(E)+y(\theta)(M)}{GNP}.$$

where, for any given year:

E = the value of all commodity exports from the thirteen colonies
x = the percentage of these exports sent directly to Britain
j = the percentage of these exports consumed in Britain
σ = the average increase in shipment costs from routing exports through Britain, expressed as a percentage of E
M = the value of all commodity imports
y = the percentage of these imports routed from foreign countries through Britain
θ = the average increase in shipment costs for imports so rerouted, expressed as a percentage of M
GNP = the gross national product of the thirteen colonies

The objective of this article is to demonstrate the unimportance of British interference. All estimation procedures are therefore directed to overstating the costs involved. Available evidence suggests that $(x-j)$ was larger than y, and σ equal to or greater than θ.[20] The above formula can accordingly be reduced to:

$$(x-j)(\sigma)\left(\frac{E+M}{GNP}\right)$$

Only four ratios are now required: x, j, σ and $(E+M)/GNP$. The first three can be estimated directly from eighteenth century data. As indicated in Appendices A and B, reasonable upper-bound approximations for $(x-j)$ and σ are 0.5 and 0.4, respectively.

The major difficulty is the absence of GNP data for the colonial period. The first detailed estimate of the ratio $(E+M)$/GNP is for the period 1834–43.[21] That estimate can still be used in the above calculation, provided that the ratio of exports plus imports to GNP in the colonial period was less than or equal to the comparable ratio in the years 1834–43. That is, one must assume:

$$\left[\frac{E+M}{GNP}\right]_{1763/75} \leqq \left[\frac{E+M}{GNP}\right]_{1834/43}=0.192$$

The problem then becomes to justify the assumption. The foreign trade of the late 1830's in many ways resembled that of colonial days. Exports continued to be dominated by agricultural staples: cotton accounting for 60 percent of the total; cotton, rice, and tobacco for 70 percent.[22]

---

[20] See Appendices A and B.

[21] Simon Kuznets, "Level and Structure of Foreign Trade: Long-Term Trends," *Economic Development and Cultural Change*, Jan., 1967, p. 113.

[22] Calculated from data for the five-year period 1837 to 1841. (U. S. Dept. of the Treasury, *Annual Statements of Commerce and Navigation*.)

The era of high protective tariffs was yet to come, as was the major thrust of industrial might with its associated pressures for import substitution. Kuznets' survey of recent trade data suggests that the ratio $(E+M)/GNP$ is positively correlated with per capita income and negatively correlated with country size.[23] What the net effect of these two opposing tendencies was for the United States between 1763–75 and 1834–43 is a matter of conjecture. Between 1834–43 and 1888–89 the American ratio was surprisingly stable, fluctuating between 0.129 and 0.117, and concluding the century at 0.125.[24] That these ratios should have been exceeded by a wide margin within the parochial setting of colonial America would seem highly unlikely. To allow for the possibility of some positive margin, the assumption is made that for the period 1763–75 the ratio $(E+M)/GNP$ was, on the average, 0.15.

The percentage of colonial gross national product sacrificed because of British interference with overseas trade can therefore be approximated by:

$$(x - j)(\sigma)\left(\frac{E + M}{GNP}\right) = (.5)(.4)(.15) = 3\%$$

What must be emphasized is that 3 percent is an upper-bound measure of the actual costs involved. The numbers used to measure $(x-j)$, σ and $(E+M)/GNP$ are undoubtedly all too large. The formula itself—even if calculated with accurate rather than exaggerated ratios—would probably overstate the "true" cost of C_2C_1KL portrayed in Figure 1. (See Appendix C.) Finally, no allowance has been made for the benefits accruing to the colonies under the Navigation Acts, notably bounties on a wide variety of exports and preferential duties on most enumerated goods.[25] Over an extended period of time, even a small loss in income can have serious implications for growth if linked adversely to such dynamic factors as capital formation, entrepreneurial activity or technological change. What the actual linkage was to such factors in the period 1763–75 remains obscure. What is not obscure is the brevity of the period during which the counterfactual hypothesis of American independence can reasonably be postulated. As long as that remains confined to thirteen years, the possibility of dynamic influences seriously magnifying the above percentage would seem remote.

A conclusion discounting the importance of income sacrificed because of British interference with colonial trade routes to Europe is hardly surprising. The geographic diversion was minor; the goods affected were only a small percentage of total American GNP. Even the colonists demonstrated a marked insensitivity to the economic burdens. The Declaration of Independence, for example, makes no mention of the Acts of Trade, and those hardest hit by export restraints—the planters of Virginia and Maryland—almost never included those restraints in their list of grievances against the mother country.[26] Two questions remain unanswered. The first is the net cost to the colonies of all imperial policies, not just the Navigation Acts. The second is the relationship between those real costs and "certain disturbances" which culminated in an independent republic. The present article is part of a larger study directed to these more general and less tractable problems of the colonial era.

---

[23] Simon Kuznets, "Level and Structure of Foreign Trade: Comparisons for Recent Years," *Economic Development and Cultural Change*, Oct., 1964, p. 25.

[24] Kuznets, "Long-Term Trends," p. 113.

[25] The net effect on colonial shipping remains obscure, mainly because of a shortage of data on American shipbuilding, sales and ownership in the pre- and post-revolutionary eras.

[26] See for example T. W. Tate, "The Coming of the Revolution in Virginia: Britain's Challenge to Virginia's Ruling Class, 1763–1776," *William and Mary Quar.*, July, 1962, pp. 324, 335.

APPENDIX A

The only comprehensive trade data for the thirteen colonies are for the five-year period 1768–72. These have been revised by James Shepherd, of Purdue University, the quantities reported in American and British customs records being multiplied by current prices.[27] According to Shepherd's data, 55 percent of all colonial exports, 1768–72, were sent to Great Britain.[28] In the period 1763–75, 87 percent of American tobacco, by weight, was reexported from Great Britain.[29] Given the dominance of tobacco in total colonial exports to Britain (about 50 percent, 1768–72) and the limited reexport trade in most other enumerated goods, 0.5 would seem to be a reasonable upper-bound estimate of $(x-j)$.

Shepherd's data show that between 1768 and 1772 only 14 percent of all American imports came from foreign countries by way of Great Britain. This ratio admittedly understates the cost of trade route distortions, insofar as the colonists also paid more for British products made competitive by incremental shipping charges incurred by European goods. What is striking is their continued reliance on British goods long after independence was achieved. In the period 1790–1800, inclusive, non-British sources accounted for only 25 percent of imports into the United States paying ad valorem duties.[30] For the period 1821–29, inclusive, when comprehensive import data for the new republic first became available, imports into the United States from sources outside the Western hemisphere other than Great Britain accounted for only 34 percent of total imports.[31] The use of $(x-j=0.5)$ for y would therefore seem to introduce a strong upward bias into the final cost estimate.

APPENDIX B

The procedure to be used to estimate σ is first to calculate σ for tobacco exports, and then to suggest why the tobacco estimate can be used for all colonial exports and imports artificially routed through Great Britain.

In the case of tobacco, what is required is:

$$\sigma = \frac{\Delta F + L}{C_1}$$

where:

C_1 = the cost of colonial tobacco, f.o.b. Maryland or Virginia
ΔF = the additional freight costs incurred from shipping tobacco to northern Europe by way of Great Britain
L = the costs incurred in British ports from landing, storing, and reloading tobacco.[32]

The only data on L that could be discovered for the eighteenth century were for three years: 1729, 1730, and 1736 (see Table 1). Cost estimates for two of these years originated with colonial planters dissatisfied with marketing arrangements of the 1730's and are unlikely to understate the expenses involved. An additional upward bias was deliberately introduced by calculating all charges per pound of tobacco on the net weight of hogsheads in

---

[27] For more details on Shepherd's correction procedure, see "A Balance of Payments for the Thirteen Colonies, 1768–1772: A Summary," *J. of Econ. Hist.*, Dec., 1965, pp. 692–93.

[28] This and all subsequent percentages referring to the years 1768–72 were calculated from data supplied to the author by Professor Shepherd.

[29] *U. S. Historical Statistics*, p. 766.

[30] Excluding 1792–94 for which data are not available. Source: U. S. Congress, *American State Papers*, VII, *Commerce and Navigation* (Washington, D.C., 1832).

[31] U. S. Treasury Department, *Annual Statements of the Commerce and Navigation of the United States*.

[32] These costs do not include payment of British duties. By 9 Geo. I, C. 21 (1722) all British import duties on colonial tobacco were recovered by the shipper when tobacco was reexported.

TABLE 1

EXTRA COSTS FROM SHIPPING COLONIAL TOBACCO TO NORTHERN EUROPE VIA GREAT BRITAIN
(Pence sterling per pound of tobacco)

	1729*	1730†	1736‡
Primage and petty charges	.030	.036	.034
Entry inwards, bonds, landwaiters	.021	.025	.025
Entry out, coquets, searchers fees	.028	.033	.033
Cooperage inwards and outwards	.028	.033	.033
Porterage, wharfage, lighterage	.028	.025	.016
Warehouse rent and cartage	.083	.041	.057
Brokerage	.028	.033	.033
Postage of letters and watching	.028	.004	.016
Freight, England to northern Europe§	.138	.145	.144
Total	.412	.375	.391

* *The Case of the Planters of Tobacco in Virginia, as Presented by Themselves* (London: J. Roberts, 1733), pp. 42, 43, 47, 48.

† Testimony of Captain John Murdock in Great Britain, House of Commons, *Report from the Committee Appointed to Inquire into the Fraud and Abuse in the Customs, 1733*, pp. 36–37.

‡ "A Memorial Relating to the Tobacco Trade," *Virginia Gazette* (Williamsburg), July 29–Aug. 5, 1737.

§ See text.

London rather than on the gross weight f.o.b. Maryland. (Intervening shrinkage was of the order of 3 percent to 8 percent.)[33]

Estimates of L in Table 1 include all payments made in Britain except a commission of 2½ percent charged for marketing the tobacco. By 1763–75, this charge and the consignment system had been substantially replaced by direct British purchases in the colonies.[34] Intensive competition, especially among Scottish factors based in Maryland and Virginia, minimized the opportunity for monopsonistic practices by marketing firms.[35] Under both systems—direct purchase or consignment—available evidence suggests that the colonists received good value for marketing charges in terms of extended credit facilities and market information.[36] Perhaps the most persuasive corroboration of the latter is the continued use of British services even after independence was achieved, over 53 percent of American tobacco exports being shipped to Great Britain in the decade of the 1790's.[37]

All sources in Table 1 gave the cost of transatlantic freight (F), but not the additional cost of moving tobacco from England to northern Europe (ΔF). Available data on the ratio $\Delta F/F$ in the eighteenth century were as follows: for American tobacco in the period 1712–18, about $12\frac{1}{2}$ percent;[38] for Barbados sugar destined for northern Europe in the 1730's, 16 percent;[39] for American rice in the 1760's, 18 percent to 30 percent.[40] The assumption was therefore made that on the average, 1729 to 1736, the ratio of $\Delta F/F$ for American tobacco was 25 percent. Reported transatlantic freight costs were multiplied by this ratio to derive the ΔF estimates in Table 1.

---

[33] Arthur P. Middleton, *Tobacco Coast* (Whittet & Shepperson, 1953), p. 104. The 1736 data, for example, are for one hogshead "weighing when shipped from Virginia 790 pounds net, and at London 732 pounds."

[34] *Ibid.*, p. 110. L. C. Gray, *Southern Agriculture*, I, pp. 425–26.

[35] L. C. Gray, *Southern Agriculture*, I, p. 433; A. P. Middleton, *op. cit.*, pp. 108–09; J. H. Soltow, "Scottish Traders in Virginia, 1750–1775," *Econ. Hist. Rev.*, Aug., 1959, pp. 90–91.

[36] See for example, L. C. Gray, "The Market Surplus Problems of Colonial Tobacco," *William and Mary Quar.*, Jan., 1928, pp. 3–4. John Lord Sheffield, *Observations on the Commerce of the American States. . . . A New Edition, Much Enlarged* (London: J. Debrett, 1784), p. 99; J. M. Sosin, "Imperial Regulation of Colonial Paper Money, 1764–1773," *Pennsylvania Mag. of Hist. and Biog.*, Aug., 1964, p. 175, n. 4.

[37] Average by weight for the decade 1790–1799. (Source: U. S. Congress, *American State Papers*, VII, *Commerce and Navigation*, Washington, D.C., 1832.) Continued use of British facilities was also apparent in rice marketing. Between 1763 and 1774, 59 percent of American rice exports by weight went to England; between 1790 and 1799, 28 percent. (Sources: *ibid.*, and *U. S. Hist. Statis.*, pp. 767–68.)

The estimation of C_1 for this period—the price of tobacco f.o.b. the colonies—was complicated by variations in tobacco quality,[41] in hogshead size, and in the exchange rate between sterling and the local currency. A survey of available sources nevertheless suggested that a price of 1 d. sterling per pound is something of a lower-bound estimate for the average price of tobacco marketed in the period 1729 to 1736.[42]

The final estimate of σ for tobacco in the 1730's is therefore:

$$\frac{\Delta F + L}{C_1} = \frac{.4}{1.0} = 40\%$$

How representative this estimate is for the prerevolutionary era depends on the trend in each of the three variables ΔF, L, and C_1 between 1729–36 and 1763–75. Transatlantic tobacco freights per hogshead remained relatively stable over this period while the average size of hogshead increased, thereby lowering shipment costs per pound.[43] Time series data on L were not available, although most of these charges were also levied per hogshead and may well have declined for similar reasons. Certainly contemporary evidence supports the impression of improved efficiency in the grading and marketing of colonial tobacco over time.[44] As for the trend in C_1, the early 1730's was one of the worst periods of the entire century for tobacco growers, in marked contrast to the relative prosperity during most of the years immediately preceding the American Revolution.[45] Between 1763 and 1772, for example, Philadelphia tobacco prices were on the average 59 percent higher than they had been in the years 1729 to 1736.[46] A comparison for the same time periods using

---

[38] J. M. Hemphill, "Freight Rates in the Maryland TobaccoTrade, 1705–1762," *Maryland Hist. Mag.*, Mar., 1959, p. 47.

[39] John Ashley, *Memoirs and Considerations Concerning the Trade and Revenue of the British Colonies in America* (London: C. Corbett, 1740), pp. 98–99.

[40] Leila Sellers, *Charleston Business on the Eve of the American Revolution* (Univ. of North Carolina Press, 1934), pp. 152–53.

[41] The lower quality colonial tobacco tended to predominate in the reexport trade to northern Europe. See A. P. Middleton, *op. cit.*, pp. 98, 129; L. C. Gray, *Southern Agriculture*, I, pp. 253–55; "A Memorial Relating to the Tobacco Trade," *Virginia Gazette* (Williamsburg), July 29–Aug. 5, 1737; Great Britain, House of Commons, *Report from the Committee Appointed to Inquire into the Fraud and Abuse in the Customs, 1733*, p. 69.

[42] For prices of tobacco in Maryland and Virginia, see Great Britain, *Journal of the Commissioners for Trade and Plantations, November 1718–December 1722*, p. 210; R. A. Brock, "A Succinct Account of Tobacco in Virginia," *Tenth Census of the United States, 1880, Agriculture*, Vol. III, p. 818; V. J. Wyckoff, "Tobacco Regulation in Colonial Maryland," *Johns Hopkins Studies in Historical and Political Science*, No. 22, 1936, p. 161; Anne Bezanson *et al.*, *Prices in Colonial Pennsylvania* (Univ. of Pennsylvania, 1935), p. 422; William W. Hening, *The Laws of Virginia from 1619* (Richmond: W. W. Gray, 1820), Vol. IV, pp. 352, 353, 422; James Bisset, *Abridgement and Collection of the Acts of Assembly of the Province of Maryland* (Philadelphia: W. Bradford, 1759), p. 159; U. S. Department of Agriculture, *Yearbook*, 1908, p. 683. For rates of exchange between sterling and local currencies, see Anne Bezanson, *op. cit.*, p. 432; William W. Hening, *op. cit.*, p. 219; James Bisset, *op. cit.*, p. 159; Hugh Jones, *The Present State of Virginia* (London: J. Clarke, 1724), p. 45.

[43] J. M. Hemphill, *op. cit.*, p. 56; Gary M. Walton, "A Measure of Productivity Change in American Colonial Shipping," *Econ. Hist. Rev.*, Aug., 1968, pp. 268–76.

[44] L. C. Gray, "Market Surplus," p. 3; A. P. Middleton, *op. cit.*, pp. 113–26.

[45] L. C. Gray, *Southern Agriculture*, I, pp. 270–75.

[46] Anne Bezanson *et al.*, *Prices in Colonial Pennsylvania*, p. 422.

Bezanson's index for twenty commodities (including tobacco) shows a general rise in Philadelphia prices of 39 percent.[47] Given this evidence of a rise in C_1, a fall in transatlantic freights and a general improvement in the marketing facilities for tobacco, the inference would seem justified that σ for colonial tobacco exports between 1763 and 1775 was no greater than 40 percent.

In the period 1768 to 1772, inclusive, tobacco accounted for 50 percent of total colonial exports to Great Britain; rice, indigo, bar and pig iron, naval stores and potash for almost 30 percent.[48] The major difficulty is the absence of data on ΔF and L for the latter commodities throughout the prerevolutionary era.[49] The use of the σ estimate of tobacco for all enumerated exports would seem defensible on two counts. With the exception of rice,[50] most enumerated goods were consumed primarily in Great Britain rather than being reexported to the continent. Equally important, freighting charges in the eighteenth century were more a function of volume than of weight,[51] tobacco being distinguished by a relatively high ratio of bulk to value. In the case of rice, for example, the ratio of transatlantic freight to American f.o.b. price was approximately half what it was for tobacco in the period 1763 to 1775.[52]

A similar data scarcity confounds direct estimation of θ for colonial imports artificially routed through Great Britain. As indicated in Appendix A, these goods constituted less than 20 percent of total imports during the period under study. British legislation permitted the colonists to buy direct a number of their principal European imports, notably salt and a variety of wines. Most imports subject to indirect routing, such as tea and pepper, had a low ratio of bulk to value relative to that for tobacco. This latter fact plus the marked exaggeration in import costs arising from the substitution of $(x-j=0.5)$ for y would seem to make defensible the substitution of 0.4 for θ.[53]

APPENDIX C

In terms of the geometry of Figure 1:

$$\sigma = \frac{\Delta T}{C_1}$$

$$(x-j)(E) = C_1Q_1$$

$$\therefore\ (x-j)(\sigma)(E) = (\Delta T)Q_1$$

The correct measure of the costs of routing colonial exports through Britain is:

$$(C_2 - C_1)Q_1 + KJL$$

Area KJL will exceed

$$(\Delta T)Q_1 - (C_2 - C_1)Q_1$$

only if the supply curve has an extreme slope relative to the slope of the demand curve.[54] Even if the slope of the latter is zero (i.e., demand is infinitely price elastic) the omitted costs included in *KJL* are trivial unless supply is extremely price elastic. For example, if demand is infinitely price elastic and supply is a linear function with a price elasticity of one, it can be shown that[55]

---

[47] *Ibid.*, p. 425.

[48] See Appendix A, footnote 28.

[49] The only other detailed cost information that could be discovered was for sugar in the 1730's, a dissatisfied Barbados planter estimating $(\Delta F+L)$ for his product at 20 percent to 40 percent of Barbados f.o.b. price. John Ashley, *Some Observations on a Direct Exportation of Sugar from the British Islands* (London: 1735), p. 4.

[50] British legislation permitted direct shipment of colonial rice to European ports south of Cape Finisterre and to American ports south of Georgia. Rice exports destined for northern Europe could be imported into England duty free, but were subject to a small export duty on reshipment (7 Geo. III, c. 47, 1767).

[51] See for example Gary Walton, *op. cit.*, pp. 272–73.

[52] Freights for tobacco may be found in Gary Walton, *op. cit.*, p. 280; for rice, in Leila Sellers, *op. cit.*, pp. 152–54. Hogshead were converted into pounds assuming for tobacco 1 hhd. = 950 lbs. (U. S. Dept. of Agric., *Yearbook*, 1908, p. 682); for rice, 1 hhd. = 2 barrels = 1050 lbs. (*U. S. Historical Statistics*, pp. 750–51). Bezanson's data on Philadelphia prices of tobacco and rice (*ibid.*, p. 772) were converted into sterling according to the rates given in *ibid.*, p. 773.

[53] Most British duties on foreign goods were repaid when those goods were reexported to the colonies. (See Harper, *op. cit.*, pp. 10–11.) Duties that were not repaid constituted only a minor expense. In 1773, for example, they amounted to only 3½ percent of "official" British values, official values in turn being an underestimate of current values. (Calculated from data in *ibid.*, pp. 28, 35, n. 90.)

$$\begin{aligned} KJL &= (0.5)(C_2 - C_1)(Q_2 - Q_1) \\ &= (0.8)(C_1Q_1) \end{aligned}$$

With comparable assumptions for imports (that supply is infinitely elastic and the demand function is linear with a price elasticity of one) it can be shown that the omitted area of consumer surplus is also $(0.8)(C_1Q_1)$. Given that a summation of C_1Q_1 for all exports and imports equals $(x-j)(E+M)$, it follows that omitted costs under the above conditions

$$= (.08)(x-j)\left(\frac{E+M}{GNP}\right) = (.08)(.5)(.15)$$

$$= 0.6\% \text{ of } GNP.$$

---

[54] If supply and demand functions are linear

$$KJL > (\Delta T)Q_1 - (C_2 - C_1)Q_1$$

only if

$$\text{slope of } S_{fob} > \frac{2Q_1}{(Q_2 - Q_1)} \cdot \quad \text{slope of } D_{fob}$$

For imports the conditions are reversed to

$$\text{slope of } D_{cif} > \frac{2Q_1}{(Q_2 - Q_1)} \cdot \quad \text{slope of } S_{cif}$$

Details on the derivation may be found in Richard A. Musgrave, *The Theory of Public Finance* (McGraw-Hill, 1959), pp. 290–91.

[55] Given:

$$\frac{\Delta C}{C_1} = \frac{\Delta T}{C_1} = 0.4$$

and

$$\begin{aligned} KJL &= (.5)(\Delta C)(\Delta Q) \\ &= (.5)\,\frac{\Delta C}{C_1}\cdot\frac{C_{1\Delta Q}}{Q_{1\Delta C}}\cdot\frac{Q_{1\Delta C}}{C_1}\cdot C_1 \end{aligned}$$

JONATHAN R. T. HUGHES: Justification for historical work of this sort is the pursuit of truth for its own sake. And for any historian, that is enough. Professor McClelland is to be commended for the work in this paper, which shows us in more pointed evidence than we had before that the Navigation Acts had no important economic consequences on the revolutionaries of Colonial America that could be measured by known evidence and quantitative techniques.

It is clear that the years 1763–75 would be the most appropriate ones for the contrafactual world of a pre-1776 independent United States. I would suggest a single emendation. Since the hypothetical world of contrafactual history need be limited only by logic and ingenuity, I would argue that the independent United States of 1763–75 (we say nothing of how it got that way) was in no danger of a military takeover from the French compared to the danger of such a takeover by the British. The French of 1763 could verify that. Any estimate of hypothetical defense expenditures to protect the fledgling Republic from a British takeover would need to be correspondingly heavier (compared to France as a potential enemy) and the net "contribution" to American national income of British colonial status would be correspondingly higher. As the Revolution and the War of 1812 showed, the British were expensive enemies.

The burden of my comments is something else. Although it is in our best interests as scholars to get these arguments pinned down as closely as possible, the question arises, was this trip necessary? I'm glad we made the trip, but we could have stayed home comfortably with Lawrence Harper.

The reason these calculations are still going on are primarily two: (1) the basic question of possible economic "causes" of the American Revolution is a popular one, and (2) Harper has been to some extent misunderstood. For example it is stated in a well-known paper:

> Lawrence Harper, in a now classic article actually attempted to calculate the cost and found that British policies "placed a heavy burden on the colonies."[1]

Harper said that, all right, but in a certain context, and that was not about "British imperial regulation" in general. Harper had a four-part analysis of the data.[2] The four parts were: (1) transatlantic trade, (2) colonial manufactures, (3) American trade, and (4) the post-1763 measures, which included the famous ones, the Quebec Act (1763 and 1774), the Sugar Act (1764), the Stamp Act (1765), the Tea Act of 1773, and the closing of the Port of Boston and the other "Intolerable Acts."

Of Part 1, and only there, did Harper make the statement quoted. Harper estimated that the burden of British policy was from $2 millions to $7 millions, depending upon whether you accepted his minimum or maximum estimates. He called this burden "heavy" by saying that it could have paid all government expenses during Washington's administration at the minimum and at the maximum could have raised in the twelve years 1790–1801 enough to pay off both the domestic and foreign debt of the U.S.A. It is this part of Harper that has been under discussion here. As for the rest, and on balance, Harper is on the other side of the fence and sees no known "economic causes" of the American Revolution.

I will now quote Harper on the rest of his points.

On Part 2 Harper says:

> . . . English mercantilism as a whole probably had a greater tendency to promote than to hinder colonial industry. (P. 102.)

On Part 3, his examination of duties on sugar, molasses, and rum, Harper says that they were "usually evaded" anyhow, and data on prices:

---

[1] Robert Paul Thomas, "A Quantitative Approach to the Study of the Effects of British Imperial Policy on Colonial Welfare: Some Preliminary Findings," *J. of Econ. Hist.*, Dec., 1965.

[2] Lawrence A. Harper, "Mercantilism and the American Revolution," *Canadian Hist. Rev.*, Mar., 1942. Reprinted in part in Nash, *op. cit.*

> . . . do not indicate that the legislation had any radical effect upon the trade. (P. 102.)

On Part 4 Harper argued that the post-1763 British measures:

> . . . aroused the resentment of the merchants who unleashed an avalanche of agitation which soon went beyond their control. The agitation was not directed toward revolution at first, but agitation by its very nature promotes conditions favourable for revolution—and revolution followed as a natural sequence. (P. 102.)

Overall, Harper says:

> . . . it is difficult to understand how British mercantilism discriminated materially against the colonists. (P. 103.)

Of the molasses taxes he said:

> The excitement aroused by these taxes leads one to look for evidence of the havoc which they wrought in trade, but an examination of the wholesale prices of molasses does not disclose any noticeable change attributable to the legislation. (P. 104.)

On the Navigation Acts Harper said further:

> The only measures which afforded a sufficient economic grievance were the *entrepôt* provisions of the Navigation Acts, which governed the trans-Atlantic trade. They helped to create a fundamental economic unbalance, but cannot be connected directly with the Revolution. The colonists had lived with them for more than a century. . . . (P. 105.)

On the "economic origins" of the Revolution in general Harper said:

> If there were any inexorable economic forces which were inevitably drawing the colonies toward revolution, they are hard to detect and the colonists were unaware of them.
>
> Anyone who maintains that the Revolution resulted from the inevitable clash of competing capitalisms must reckon with several points: That burdens upon the trans-Atlantic trade were proportionately greater at the beginning of the eighteenth century than in 1776; that the restraints of the land and currency policies were basically the same as those prescribed by the federal government [i.e., after independence]; and that after 1766 the taxes laid on molasses by Britain were less than those imposed by the United States after 1790. (P. 65.)

I submit that to imply that Lawrence Harper somehow argued that British mercantile policy in general was burdensome and hence contributed to the American Revolution in a measurable way is a curious position indeed.

Yet there was a revolution and the revolutionists complained long and loudly of oppression by the King in terms of economic factors. If I may quote from the Declaration of Independence for a moment:

> He has endeavoured to prevent the population of these States; for that purpose obstructing the Laws for Naturalization of Foreigners; refusing the pass others to encourage their migrations hither, and raising the conditions of new Appropriations of Lands.
>
> He has erected a multitude of New Offices, and sent hither swarms of Officers to harrass our people, and eat out their substance.
>
> He has combined with others. . . .
>
> For cutting off our trade with all parts of the world. . . .
>
> For imposing Taxes on us without our Consent. . . .

As you know, other "economic" complaints are found easily enough in the Declaration. But here alone the colonists complain of restraints upon land, population, the high cost of a foreign bureaucracy, constraints upon trade, unjust taxation. In a way it doesn't make any difference whether or not such complaints were justified, so long as the colonists were encouraged to raise complaint to military violence thereby. We know, for example, that in September, 1765, the newly designated Stamp Tax Officer of Philadelphia (a representative of one of the "oldest" families and hence no "foreigner") was "waited upon" by a "deputation of the mob with muffled drums and muffled church bells ringing."[3] Since the Stamp Tax Officer in question had not yet received either his commission or his stamps, the mob could not have been on his doorstep because they were being oppressed by the new tax. Then why were they there?

This brings us back full circle to the question about the Revolution. Were there any real "economic" origins? The older members of this audience will have long since realized that a ghost has been lurking behind Harper's enormous research effort and this whole discussion. The ghost, of course, is Louis Hacker's famous 1935 paper (expanded in *The Triumph of American Capitalism* in 1940), "The First American Revolution," in which he came down squarely on the side of economic causation. But not of the crude sort represented by data. Extracted from the rhetoric of 1930's radicalism one is struck today by several aspects of Hacker's thesis. First, as Harper said, there was nothing novel about the idea that American businessmen were more or less fed up with British regulation. But Hacker said more than that. He claimed that it was inevitable in the laws of the British mercantile system that the colonial business interests had no reason to suppose by 1776 that they could ever get control of their own affairs. The British mercantile system had never changed from its seventeenth century origins. Control of the Empire's business interests was to remain in London. That was the importance of the Tea Act:

> . . . that not only was the economic vassalage of the American colonies to be continued but the interest of colonial enterprisers was to be subordinated to every British capitalist group that could gain the ear of Parliament.[4]

Hence, according to Hacker, they were glad to back down on specific taxes if the principle of Parliamentary control remained. It was that principle which could not be negotiated. The consequences were both economic and political—not one or the other.

But by 1776, when independence was declared, revolutionary violence, provocation and suppression had become a settled part of the American scene, beginning with the Boston Massacre in 1770 and escalating to the eruptions of 1775. The Port of Boston was closed June 1, 1774, and that alone by the standards of today's hotheaded statesmen would have been a *causus belli*. Part of the causes of any of this was economic. To say that all of it was economic would be the height of folly, and no reputable historian ever did say that. To be able to say how much of the revolutionary motivation was in terms of shillings and pence per head would be a work of inspired genius. We are not the recipients of such a creation yet. I don't think we are headed that way. I think it is good that some very fine talent and hard work has demonstrated that the quantification of the economic influences behind the revolutionary militants of the 1770's lies beyond our grasp at present. Here I commend McClelland for this effort.

HERMAN E. KROOSS: In this era when the search for the magic number has become the Holy Grail of economic history, McClelland's paper is refreshing. It is short. It avoids the worst sophistries of Occam's Razor and resorts to the fundamentals of simple economics. The methodology and logic are impeccable; the conclusion succinct and irrefutable: the upper-bound cost of the Navigation Acts' trade route distortions were 3 percent of colonial GNP.

But McClelland's paper also illustrates the depression that some of us feel when we watch the game of "Find the Right Number." Thomas, whose "conceptual framework . . . is so defective that [his] final result is meaningless," concluded that the cost of an American being an Englishman was $1.24 in 1770. Had he used Spanish money, he might have found it to be $1.24 1/12. I am reminded of another such specific number: Chase's estimate of the initial cost of the Civil War, $318,519,581.87. Harper, contemplating much more than just the Navigation Acts, estimates the "heavy burden" at between $2 million and $7 million.

---

[3] George Bancroft, *History of the United States of America*, 1879 ed. (Little, Brown), Vol. III.

[4] Louis M. Hacker, "The First American Revolution," *Columbia Univ. Quar.*, Part 1, Sept., 1935; reprinted in part in Gerald D. Nash, *Issues in American Economic History* (D.C. Heath & Co., 1964).

Given the difference in methodology, by how much does McClelland's conclusion differ from that of Thomas and Harper? Assuming a GNP of $100,000,000, Thomas' estimate of the direct costs of the Navigation Acts is 2.7 of GNP, compared with 3 percent; $1.20 per capita as against $1.40.

What is really vexatious about McClelland's paper is the topic. The cost of the Navigation Acts has become a great bore. No matter how much one beats an unlive horse, it still remains dead. Since before Beer and after Harper, historians and economists of great talent have spent time, talent, and travail on "reassessing" the previous generation's interpretations.

And there really hasn't been that much difference. Elementary economic theory, which abhors a free lunch, long since led us to the conclusion that the Navigation Acts by themselves and *sans* reference to imperial protection cost the colonists something. It could not have been much because the colonists didn't have much. How much, then, is much? According to McClelland's estimate, the cost of the Navigation Acts was exactly the same percent of GNP that the Vietnam action is now costing the United States, which is or is not very much, depending upon where the observer sits in judgment.

Disagreements among historians involve much more than the dollar-and-cents direct costs of the Navigation Acts. The "cement of empire" and the "ubiquity of smuggling" are the issues that remain arguable. As McClelland says, "What remains obscure is the motivation for colonial indignation." And precise methodology and the right numbers do not cast any further light on this essential obscurity.

ECONOMIC DEVELOPMENT IN HISTORICAL PERSPECTIVE

By JOHN C. H. FEI, *Cornell University*
and
GUSTAV RANIS, *Yale University*

It is the purpose of the present paper to discuss the process of economic growth in long-run historical perspective, in the belief that an appraisal of the evolution of crucial economic institutions and significant growth theoretic ideas may help illuminate the state of our present-day understanding of growth. In order to carry out this purpose, we must, first of all, decide on the relevant time span attached to our notion of recorded history. As we are concerned, in part, with the evolution of economic ideas about growth, the history of economic thought immediately suggests that the time span of our inquiry should stretch back to at least the sixteenth century. For it is in respect to this period that economists have expressed themselves in an organized body of thought (i.e., the mercantilists, the physiocrats, the English classical writers, the Marxists, the Austrians and contemporary writers), either explicitly or implicitly directed to the phenomenon of growth. An examination of these ideas provides one of the major building blocks of this paper.

Growth is admittedly a complicated phenomenon which can be interpreted strictly narrowly in terms of pure resources augmentation or, more broadly, as a complete cultural evolution. We have found it instructive to adopt a viewpoint which lies somewhere in between these extremes; i.e., by concentrating on broadly defined notions of capital and the capital accumulation process over time (Section I). To be sure, the productive significance of capital, *à la* Harrod-Domar, has remained important; but the capital stock is also an instrument of control and intimately related to the organizational features of an economy. If we take a sweeping view of the growth accomplishments of the Western world over the last five centuries (from 1500 on), we cannot help but be impressed by the drastic changes, not only in the productive significance of capital, but also the mode of social organization in different growth epochs.

This basic notion of a growth epoch, as Professor Kuznets has taught us,[1] can be defined in terms of certain unifying rules of growth or modes of operation of the whole economy over a rather long (at least one hundred years) stretch of time. The epochs we shall be dealing with are, first, the premodern epoch of agrarian feudalism yielding to mercantile agrarianism during the period from approximately 1500 to 1750. This is followed by what Kuznets has identified as the epoch of modern growth and which can be further subdivided into nineteenth century industrial capitalism (1750–1914) and twentieth century technocratic capitalism (1918–present). For each epoch, we shall try (Sections II-III) to depict the special relevant meaning of "capital," of "capital accumulation," of the organizational structure of the economy, and of pertinent growth theoretic ideas. The focal point of our analysis will be on the contrast between the various epochs as well as on the significance of the sequential ordering outlined; namely, in what sense the growth accomplishment of one epoch paves the way for the next.

---

[1] See S. Kuznets, *Modern Economic Growth* (Yale Univ. Press, 1966).

The contemporary less developed world is trying, in the course of a few decades, to imitate Western European experience with growth over the last four centuries—both in terms of its productive and its organizational aspects. The resistance that is likely to be encountered in this attempt to telescope historical experience has substantial policy significance for the development of the contemporary underdeveloped world and will be explored in Section IV.

I. *Capital Accumulation*

Capital is generally defined as a stock which represents a produced means of production. Three essential attributes (and only three) can be deduced from this; namely, that the accumulation of capital requires social effort (i.e., it is "produced"), that it is an immutable and durable stock, and that it contributes to production. This definition permits a broad interpretation of what may be included in the capital stock; i.e., inventories, fixed capital (e.g., plant and productive equipment), overhead capital (roads, wharves, water supply, schools, sewage systems, dwellings), and special quality characteristics of human agents above the unskilled level (i.e., the physical and mental attributes of labor, entrepreneurial ability, etc.). As the mode of operation of the economy changes over time, the content of "capital," within the above broad definition, undergoes a corresponding evolution.

This definition of the content of capital is quite independent of the mode of social organization. In the literature, there is, however, something of a tradition by which the term is restricted to profit-seeking private capital in a capitalistic society. This special usage serves to emphasize the fact that capital, besides constituting an instrument for raising productivity, in the engineering sense, also constitutes an instrument of control in a particular type of society; i.e., capitalism. It serves us well in explaining the growth-promoting forces of industrial capital that prevailed in that particular historical stretch of time; i.e., from 1750 to 1914. For it was in this phase that economic individualism found its fullest expression in the accumulation and management of the most characteristic form of capital (i.e., fixed capital) vital for the growth of the industrial economy. We should not let this blind us, however, to the realization that, given a longer historical time perspective, such usage remains rather special. For in contrast to economic individualism, collectivism may have had a more essential guiding role in the management of economic affairs at other times; i.e., before 1750 or after 1914. In such epochs profit-seeking private capital may have been less important in terms of economic growth than a collective form of capital management and accumulation.

Both the production and the organizational significance of capital derive from its involvement with labor. The production significance of capital is due mainly to the "contribution" which it makes to labor productivity; and the organizational significance of capital centers on the quality of this collaboration. In fact, as the economy moves from one growth epoch to another, its mode of operation is defined mainly in terms of these changing relations between labor and capital.[2] In each epoch of growth the center of the stage is held by certain special types of capital goods

[2] This view of capital as inextricably intertwined with the special growth phenomenon of an epoch is, by no means, shared by all who concern themselves with capital theory. The Austrian economists, for example, whose technical contribution to the "nature of capital" was very significant, paid little attention to the specific nature of capital in the context of a particular mode of organization of the system.

and certain characteristic modes of productive as well as organizational relationships which are evolved to facilitate the accumulation process. An understanding of this is essential to an understanding of growth in long-run historical perspective.

II. *From Simple Agrarianism to Mercantile Agrarianism (1500–1750)*

Before 1500 Western Europe was characterized by local self-sufficiency in agricultural production, or simple agrarianism. The 250 year time span from approximately 1500 to 1750 witnessed the gradual transition of this system to a less primitive, trade-related mercantile agrarianism. Mercantile agrarianism in turn paved the way for the epoch of industrial capitalism (1750–1914) which followed. Thus, in terms of our attempt at achieving long-run historical perspective, mercantile agrarianism may be regarded as a long phase of transition leading from simple agrarianism "inevitably" to industrial capitalism.

Simple Agrarianism. The simple agrarian society is characterized by the dominance of settled (non-nomadic) agricultural production to the virtual exclusion of other forms of economic activity and with relatively little trade beyond the local community. A simple agrarian society is essentially locally self-sufficient and should not be associated with the notion of an integrated national or even regional economic system. The political structure associated with such an inward-looking economic way of life was characterized as local separatism stabilized by a feudal dichotomy between hereditary ruling and serf classes. The well-known manorial system of medieval Europe and of the (less well-known) Chou dynasty in ancient China (100 to 200 B.C.) may be viewed as representative.

The nature of the capital stock in simple agrarianism represents the purest form of "wages fund"; namely, the stock amounts to an inventory of agricultural goods (mainly food) to bridge the gap arising from the non-coincidence of production and consumption periods in agriculture. On the one hand, the seasonality of agricultural production leads to the emergence of the required food supplies during one or two specific (harvest) months of the year. On the other hand, the consumption demand for food is continuous and evenly spread throughout the year. Thus all the food needed between harvests must be stocked up and social organization devised to ensure that such stocks will be apportioned and again replenished in an orderly fashion. This, in essence, is the meaning of capital stock—and the only possible meaning—in the simple agrarian society.

The mode of economic life in simple agrarianism may be depicted with the aid of Figure 1a in the form of one production sector and two household sectors; i.e., the serfs and the nobility. Total output A is used either as consumption (for the serfs C_s and for the lords C_L) or as investment (I) leading to capital accumulation. The stock of capital K is seen to be divided into twelve parts—corresponding to the evenly distributed consumption demand during each of the twelve months of the year—to remind us of the basic wage fund characteristic. While the serfs supply the necessary agricultural labor force (L), the lords supply their services in respect to the management of the capital stock K, as a wages fund.[3] The right of the lords to management is maintained and perpetuated partly by ideology (e.g., religious or feudal), partly by brute force, but mainly by the necessity for all to accept some form of social organization to ensure order. In fact, the ***raison d'être*** of

---

[3] The lords have the right to exact services and payments in kind and are in charge of settling disputes, maintaining justice, granting loans, etc.

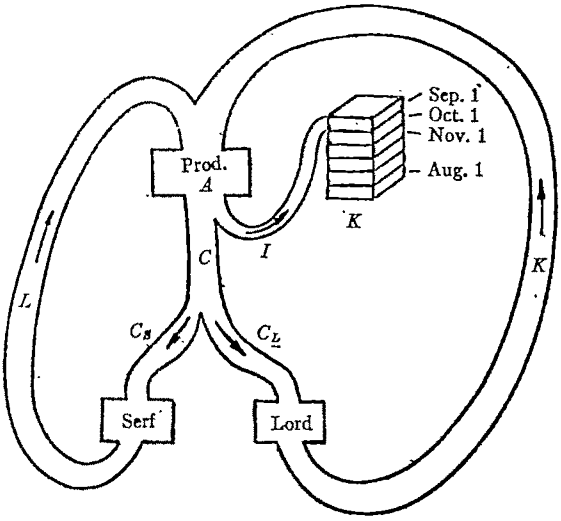


FIGURE 1a

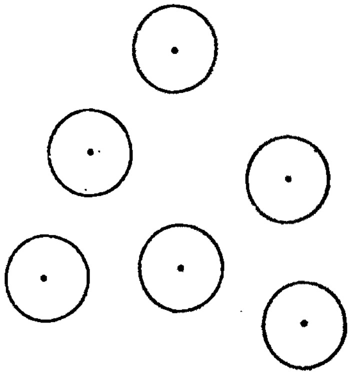


FIGURE 1b

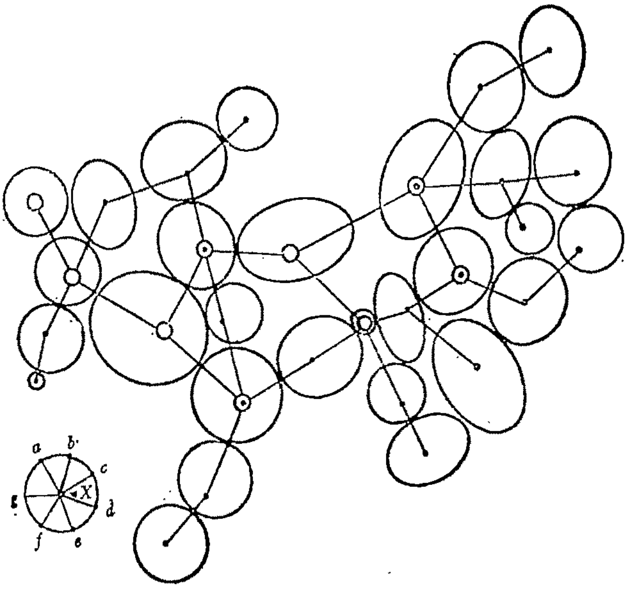


FIGURE 2a

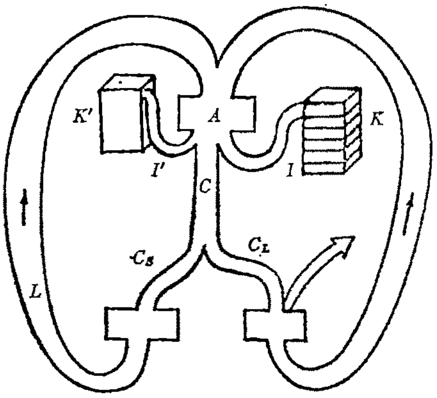


FIGURE 2b

simple agrarianism, which contributes to its long-run stability, often rests on the ground that the cultural life of the lords is taken as the very end purpose of the existence of society. This was the picture presented by Quesnay and the physiocrats.

The nature of the involvement of capital with labor in simple agrarianism is mainly to feed the workers in anticipation of the next harvest. This explains the central fact that the capital stock (K) is proportional to the population (L) where the proportionality factor Θ in $K = \Theta L$ depends upon and is positively related to the "degree" of divergence between the production and consumption periods.[4] It follows that investment per head (I/L) in such a society must be proportional to the population growth rate (n_L) with the same proportionality factor; i.e.,

$$I/L = \frac{dK}{dt} \Big/ L = \theta n_L. \tag{1}$$

For example, in case the population growth rate is constant the investment per head is also constant. Thus with population constituting the major investment demand, growth can be said to be population pushed. Where land supply is not a constraint the tasks associated with "savings" are automatic and performed routinely and almost subconsciously. Any sense of economic progress beyond this simple sideward motion is not only foreign but basically repugnant.[5]

Mercantile Agrarianism. Change and progress in the agrarian society can be closely identified with the increasing impact of mercantile activities; i.e., interregional and international economic arbitrage carried out for the sake of profits. It is the growth in the volume of trade and its geographic spread which constitutes the central growth phenomenon of mercantile agrarianism (approximately 1500–1750). To be sure, the major form of economic production—measured in terms of value added or population involved—continues to be agriculture. However, it is the penetration of this agrarian system by mercantile activities dedicated to interregional trade that gradually transforms the latter—ridding it of the local self-sufficiency attributes and substituting in its place an integrated economic system covering a larger space, ultimately the entire national economy.

What sharply differentiates mercantile agrarianism from simple agrarianism is the newly erected social infrastructure pertinent to this particular type of trade-related space economics. The structure in evidence by the end of the mercantile agrarian epoch can be depicted symbolically by a tree-star structure,[6] in Figure 2a. One set of ideas conveyed refers to the nodes or vertices (represented by O, o, .) and the edges (represented by links) connecting the nodes. While the nodes stand for communities of human settlement—O for city, o for town and . for village—the edges stand for transportation and communication links connecting these communities. The second set of ideas refers to the stars, represented by the circles in Figure 2a, surrounding the nodes, which may be construed to represent a geographic area of agrarian activity. In these areas economic activities are carried out around the center of gravity at the node or the center of the star. Thus while the trees (i.e., the nodes and the edges) signify the existence of a significant interregional pattern of connectivity, it is the stars that carry the spacial significance with respect to given geographic areas. The means of local transportation can be assumed to be fairly primitive, and hence the maximum distance between the center of the star and any point inside the star is such that it takes at most a day to complete a round trip by walking. In a densely populated country the inhabitable land space can be considered to be covered by such stars.

---

[4] For example, if there are two crops a year instead of one, the needed capital stock is halved. The determination of the magnitude of Θ is analogous to that of the modern transaction velocity of money.

[5] When land scarcity is a factor some slow improvement of crop practices or innovation must occur to offset diminishing returns. The authors deal with this situation in their "Agrarianism, Dualism and Economic Development" in *The Theory and Design of Economic Development*, edited by I. Adelman and E. Thorbecke (Johns Hopkins Press, 1966).

[6] Both these terms are borrowed from linear graph theory and are used in a heuristic and nonrigorous fashion here. A rigorous definition of a tree is a linear graph which is connected and circle free (i.e., free of loops) and signifies that there is one and only one path between any two nodes. Thus rigorously a tree corresponds to a primitive transportation system linking all the cities. A star is a number of nodes (e.g., *a,b,c,d,e,f,g* in Figure 2a) which are connected by one edge each to a vertex (e.g., *X*) which is the center of the star. Thus the nodes of the stars correspond to rural families and the center to the village which is the focal point of activity of a number of rural families.

In order to contrast this mercantile agrarian system sharply with simple agrarianism, we also present in Figure 1b a graphic description of the latter. In this diagram there are only stars; the absence of a meaningful tree structure is intended to convey the locally self-sufficient nature of the agrarian economy.[7] Thus, it is the establishment of the trees—i.e., the trading routes—which transforms the simple agrarian economy into a regionally or nationally integrated economy.

Mode of Operation of Mercantile Agrarianism. The basic structural characteristic of mercantile agrarianism is that the economy now moves away from its monolithic emphasis on agricultural production. It is the coexistence of agricultural and nonagricultural ways of life which now becomes its basic structural characteristic. For simplicity, in relation to our tree-star structure, we may think of the cities and the towns as corresponding to the trade sector, as it emerges out of increasingly regularized regional agricultural markets and trade fairs, with the small villages, which constitute the center of the stars, representing local agricultural communities. In conformity with this emerging new trade sector is the emergence of a new capital concept which now includes inventories of food (K), which serve as a wages fund, and inventories of agricultural produce, possibly semiprocessed (K'), which together make up what may be called the "commercial capital stock."

The mode of operation of the mercantile agrarian economy at any given time can now be described in Figure 2b, a slight modification of Figure 1a. Total agricultural output A is now seen to be divided into three parts: consumption C, investment in the wages fund I, and investment in commercial capital I'. These two types of investment lead to the augmentation of the wages fund K and the commercial capital stock K'.

The wages fund (K) continues to bridge the gap between the production and the consumption periods in agricultural output, as in simple agrarianism. The appearance now of a need for commercial capital (K') is due to two factors; namely, the non-coincidence of production and consumption periods among producing areas, and the time consumed in transport. Given these basic factors, the demand for this form of capital (K') is proportional to the volume of trade (T) (i.e., $K' = \Theta' T$).[8] Let us assume that the volume of interregional trade (T) in such an economy is mainly a function of the size of the "trade margin"; i.e., that portion of total output which is not self-consumed. Then $T = Lq$ where L is total population and q is the per capita trade margin. Let us approximate q by $q = p\text{-}c$ where p is the average labor productivity and c is the per capita consumption of self-consumed goods. Then $K' = \Theta' L (p\text{-}c)$. From this we can see that per capita investment in commercial capital $I'/L = \Theta'(p\text{-}c)\ \eta_L$ where η_L is the population growth rate. If we add the traditional demand for capital as a wages fund to this (see (1)), the total commercial investment per head required becomes

---

[7] This use of stars to describe a simple agrarian society emphasizes not only its relative backwardness but also its rather advanced state relative to really primitive nomadic societies where even such geographic focal points for permanent community activity do not exist.

[8] Where Θ' is the factor of proportionality. For example, other things being equal, the demand for commercial capital decreases when the needed transport time decreases or another region with more complementary production pattern enters into the trading orbit. Both of these events would be reflected in a decline of Θ'.

$$(2) \qquad I/L = (\theta + \theta'(p - c))\eta_L$$

Simple as this formula might be it does serve to emphasize two distinct ideas related to the growth promoting forces in mercantile agrarianism. In the first place, growth is population pulled; i.e., investment per head is proportional to the growth rate of the population (η_L) as in the case of simple agrarianism.[9] But, in the second place, the growth promoting forces are now also determined by—i.e., proportional to—the size of the agricultural trade margin (p-c).[10] Thus it would appear that in a mercantile agrarian system economic growth is both population pulled, as before, and agricultural productivity pushed.

Cumulative Growth Under Mercantile Agrarianism. At this point let us advance the hypothesis that there is interaction between I' and p; i.e., that the accumulation of commercial capital in turn contributes to the increase of agricultural productivity. Under this hypothesis, commercial capital accumulation will in turn result in an increase of the trade margin (p-c).[11] This increase will in turn lead to a faster rate of capital accumulation and thus reinforce the p-raising process. In this way we can depict a process of cumulative growth in the mercantile agrarian system. It should be emphasized that these forces making for continuous increases in agricultural productivity are new to the agrarian economy and contrast sharply with the picture of long-run stagnation gripping the simple agrarian system. To provide a slightly more rigorous formulation of this argument, let us assume that the increase of agricultural productivity is proportional to the change in the commercial capital stock per head, with j, the factor of proportionality, defined as the productivity enhancement coefficient:

$$(3) \quad dp/dt = jI/L \text{ implying, by (2) that}$$

$$(4) \quad a) \ \frac{dp}{dt} = A + Bp$$

$$\text{where } A = j(\theta - \theta' c)\eta_L$$

$$\text{and } B = j\theta'\eta_L$$

$$b) \ \eta_p = A/p + B$$

$$c) \ \eta_p = B \quad (\text{for } p \rightarrow \infty)$$

Thus in the long run the rate of increase of agricultural productivity will always take on a positive value $B = j\theta'\eta_L$. Furthermore, the long-run rate of progress is seen to be directly proportional to j, the enhancement coefficient, which describes the extent to which the accumulation of commercial capital serves to stimulate agricultural productivity. We also see that a higher population growth rate continues to be beneficial in stimulating agricultural productivity. This characterization surely is not far from the truth for pre-industrial growth in which population growth and the agricultural productivity-raising effect of markets represent the two primary growth-promoting forces.

The above may be regarded as our central hypothesis on the historical significance of the epoch of mercantile agrarianism; namely, that it was mainly through the expansion of trade that agricultural productivity was raised and the tendency to stagnation reversed. This stimulation was, in fact, so strong that, given the benefit of historical hindsight, it rendered the classical pessimism, based on the shortage of land and diminishing returns, quite irrelevant and artificial. Empirically, we have, in fact, witnessed an agricultural revolution during the period of mercantile agrarianism (1500–1750) which preceded the industrial revolution. Our line of argument in attempting to interpret this real world phenomenon depends, of course, on the strength of the behavioral assumption related to j; namely, on how effectively commercial activities can, in fact, stimulate agricultural productivity. It is to this problem that we will now turn.

---

[9] As long as there continues to be no problem of diminishing returns on the land.

[10] In a predominantly self-sufficient economy p-c is close to zero. As agricultural productivity increases, the trade margin is likely to increase, not only absolutely, but also as a fraction of p, with consumer preferences becoming more diversified with higher income levels. Thus, $(p\text{-}c)/p$ is an increasing function of p (or c/p is a decreasing function of p). In the formulation in the next section we shall approximate this phenomenon by the assumption that c is approximately constant as p increases through time.

[11] Especially with c constant as we have assumed (if c should rise somewhat, it will surely rise less than p).

Trade and Agricultural Productivity. As we pointed out earlier, when mercantile agrarianism is compared with simple agrarianism the most striking difference is the appearance of the tree-star structure linking up local economies as part of an integrated regional and, ultimately, national economic system. The immediate tangible effect of such an integration process is that, for a number of reasons, agricultural productivity is stimulated.

1. *The Smithian Division of Labor.* A forceful and well-known argument relevant here was offered by Adam Smith: first, farm productivity is enhanced by an increased division of labor; second, an increased division of labor is rendered possible through greater exchange and trade; and, finally, increased trade results from the accumulation of commercial capital. Referring to Figure 2a we can readily see the significance of our tree-star structure as a catalyst in facilitating interregional production specialization and the flow of commodities. The construction of roads, turnpikes, and communication links constitute the tree branches through which trade flows, and the provision of social overheads at the center, e.g., warehouses and financial facilities, make the star system possible.

We believe that the Smithian body of thought (and the old classical system in general) was directed precisely to a description of this mercantile agrarian society.[12] His "capital" was basically a synthesis of a wages fund and commercial capital, with the common purpose of expanding the division of labor through trade. In fact, the classical pessimism stems from the very notion that the sources of productivity increase are traceable mainly to the "division of labor," making it inevitable that diminishing returns, coupled with the shortage of land, could sooner or later be expected to win out.

2. *Changes in the Method of Organization.* The stability of the simple agrarian system is ensured by the feudalistic rights of one class of economic agents (i.e., the nobility) to control another (i.e., the serfs). In such a system, the privilege to own and manage capital goods—i.e., the agrarian wage fund—is merely a "symptom" of more fundamental hereditary rights. The coming of mercantile agrarianism, however, heralded a new method of organization; namely, a capitalistic approach which at first coexisted side by side with the old but, as it proved its superior economic efficiency, gradually replaced it.

This thesis of the evolution of economic institutions, to which Marx made a major contribution, belongs properly to the domain of the modern theory of organization, dealing with matters of incentives, coordination, authority, and information. With the emergence of mercantile agrarianism the very right to organize and control the economic affairs of the nation comes to rest increasingly in the hands of those who own and control the commercial capital stock rather than those of the landed aristocracy. The distinguishing organizational feature of the new system resides in its basic task-oriented approach according to which clearly defined economic tasks—e.g., production, management, bookkeeping—are assigned to various groups of economic agents along functional lines—and rewards bestowed accordingly. These built-in incentive cum coordinating devices of capitalism are highly conducive to economic progress, partly because of the encouragement given to individual initiative and partly because of the reinforcement provided by the social prestige attached to the ownership of capital and the power to control which it conveys.

---

[12] While Smith certainly considered fixed industrial capital, this does not loom large in his analytical explanatory apparatus. As Schumpeter put it: "The manufacturing industry that economists beheld and reasoned about was all along the industry of the artisan." (*History of Economic Analysis*, New York, 1954, p. 386.)

This new "bourgeois" method of organization was first tried and experimented with in connection with the incipient growth of trade and commercial activities. However, with the emergence of the tree and star constellation, the method spread to agricultural production. The well-known tripartite division of labor, *à la* Smith, in fact depicts this as the dominant model of British rural economic organization, with capitalist farmers renting land from the nobility and hiring free labor. This acceptance of the commercial capitalistic form of organization in agriculture may be viewed as among the most remarkable growth accomplishments of the period.

3. *Learning by Contact.* The star structure of mercantile agrarianism denotes a pattern of rural life revolving increasingly about the nucleus of the community; i.e., the villages or small market towns which comprise the center of the stars. Such village centers offer, beyond the exchange of products, an opportunity for the exchange of ideas through growing human contact and the ever expanding vision of alternative ways of economic and cultural life. It is through such contacts and ideas that new agricultural practices are devised and new combinations of traditional and nontraditional inputs tested. The tree structure likewise not only serves to facilitate the movement of goods but also the movement of ideas. The significance of the latter in explaining the agricultural revolution is that it is not only essential for that contact among men which induces inventions but also for the spread of the resulting innovations across land space. The significance of this process is underlined by students of both Western European development[13] and of Tokugawa Japan.[14] Historically, an agricultural revolution is usually achieved more via the transmission of best technology; i.e., narrowing the gap between average and best practice within a country rather than by the continuing improvement of best technology. The logical force of this argument rests on the grounds that inventions by their very nature are likely to occur in isolated pockets of progress,[15] and that a spreading and dissemination process is essential if a revolution of the agricultural economy is to be accomplished. In short, sustained agricultural progress is heavily dependent on the extent to which the agricultural sector is linked up and involved with the trading activities outside of that sector.

4. *Emergence of a National Economy.* A final set of factors strongly contributing to the strength of *j* relates to what may loosely be called incipient nationalism. Increasing class mobility, tied in with the beginnings of an egalitarian spirit and the Protestant Ethic, strengthened the forces of national trade expansion and commercialization emanating at the local level; and the emergence of the first real na-

---

[13] W. A. Cole, *British Economic Growth, 1688–1959,* with Phyllis Deane (Cambridge Univ. Press, 1962).

[14] T. C. Smith, *Agrarian Origins of Modern Japan* (Stanford Univ. Press, 1959).

[15] Either as a consequence of purely private or social (government) research effort.

tional consciousness in replacement of local and regional loyalties provided the necessary ideological cement for commerce following the new-found flag. Local trade barriers were progressively lowered, the sanctity of contract and of private property recognized. National transport and communication networks were constructed and a national currency and national financial institutions created. In other words, a viable tree structure was fashioned from a combination of physical overheads and legal-cum-institutional change.

This continuing movement away from self-sufficient and inward-looking simple agrarianism and towards an interdependent outward-looking mercantile agrarianism did not, of course, stop at the boundaries of the newly emerging nation-states. There was trade among the countries of Western Europe; and beyond that the discoveries in the New World and the resulting inflow of species accelerated competition for trade and territory and gave a further substantial fillip to the forces of Western European commercialism.

In summary, it was this increasing spread of commercial activity which marked the gradual transition from simple to mercantile agrarianism. The growing importance of commercial capital occasioned by the requirements of expanded trade and an enhanced division of labor in turn levered agricultural productivity increases and thus capital formation at its source. Moreover, these cumulative ever expanding forces inevitably led the system towards transition to full-fledged industrial capitalism. We shall turn to this subject in the next section.

III. *From Mercantile Agrarianism to Industrial Capitalism*

Western Europe's transition from mercantile agrarianism to industrial capitalism occurred in the course of the eighteenth and nineteenth centuries. This transition marks a most important landmark in human history, for it indicates the end of an agrarian age which stretched all the way back to the beginnings of civilization when the domestication of vegetables and animals first became a predominant mode of production. With the industrial revolution the epoch of modern growth was launched, characterized by rapid structural change and unprecedented sustained increases of labor productivity and per capita income.

This dramatic change, however, was far from sudden. For the development of mercantile agrarianism had prepared the way and planted the seeds for the new epoch which followed. Z goods production[16] and textile industries existed long before the industrial revolution took hold. But all we can hope to do here is try to capture the essence of major changes in the mainstream of a society in transition. Let us now turn then to a somewhat more precise description of the nature of this transformation to industrial capitalism and then proceed to a discussion of the intrinsic nature of capital, of the method of organization, and of the nature of the growth promoting forces characteristic of that epoch.

Establishment of Dualism. The most striking change in the economic landscape with the arrival of modern industrial growth is the establishment of a new form of economic dualism, characterized by the coexistence of agricultural and industrial production activities. This dualism differs from the earlier agrarian-trade "dualism" in that there now exists for the first time a large class of peasants freed from feudalistic bondage and constituting an urban labor force. Moreover, these workers now collaborate with industrial fixed capital

---

[16] Stephen Hymer and Stephen Resnick, "A Model of An Agrarian Economy Including Non-agricultural Activities" (to be published in the *A.E.R.* June, 1969).

—plant and equipment—which takes on a productive significance far exceeding that of "mobilizing labor" which was the trademark of commercial capital. We have a new central mode of production as well as a new mode of economic organization.

The contrast in the economic life of rural agricultural and urban industrial production is, in the first place, a contrast in space economics. While the extent of the feasible division of labor in agriculture is restricted by "nature," production in the city is increasingly a matter of human endeavor and ingenuity, either embodied in the fixed capital designed to cooperate with the labor force—or in changes in the quality of that labor force achieved through formal or informal learning processes.

This new dualism came as a natural outgrowth of mercantile agrarianism. The push of expanding agricultural productivity permitted a declining (rural) fraction of the total population to supply the growing urban class with its physical sustenance. Thus, there emerged for the first time in any really major sense the phenomenon of an agricultural surplus. It was this rapid expansion of agricultural productivity and the accompanying possibilities for major labor reallocation that lies at the heart of the transition from mercantile agrarianism to industrial capitalism.

Side by side with the push of agricultural productivity increase is the pull emanating from the cities. The very existence of the cities (and the roads leading to them) holds out the promise of a new physical environment and a new way of life to the rural population. Moreover, arrangements were made to channelize the agricultural surplus so that new employment opportunities were provided by the growth of a fixed capital stock which pulled the labor into the cities. Not infrequently these arrangements may be nonmarket (or institutional) in nature, including kinship donation, private charity, or public and church relief. However, more and more such arrangements are replaced by the transfer—through the market or the government's fiscal power—of the agricultural surplus for productive purposes; i.e., as wage goods for the newly employed urban workers. Thus the reallocation of labor, the channelization of the agricultural surplus to the industrial sector and the accumulation of industrial fixed capital stock together constitute the basic phenomena in the operation of early industrial capitalism.

Nature of Capital in Industrial Capitalism. As mercantile agrarianism gave way to industrial capitalism the nature of capital changed in respect to both its productive content and its organizational significance. With respect to content, it shifted gradually from the circulating variety which still dominated the classical system (1780–1830) to fixed, with which Marx was primarily concerned. The fact that Marx made a valiant effort to incorporate both types of capital in his analysis of growth—and the fact that he was practically the first and the last economist of any stature to do so—testifies to the transitional nature of the work of Marx. After him economists began to be preoccupied almost exclusively with fixed capital as a permanent and dependable source of productivity gain, due to changes in the quantity as well as the quality or scientific knowledge embodied.

With respect to the method of organization, Marx also made the point that, now for the first time, capital also becomes a primary instrument of control over a large proletarian class. However, Marx did not correctly anticipate the evolution of the capitalist system over time, including: the growth of unionism and the increasing political power of labor in a democratic setting; the increasing separation of the ownership and control of capital goods leading to the emergence of a new managerial class; and the trend toward increased government participation in decision making, and the emergence of a new class of public agents.

Alongside the changing class structure is the increasing importance of the allocation of resources to educational (i.e., investment in human capital) ends and the institutionalization of research, both public and private. As a result, the search for applied scientific knowledge has been substituted for capital augmentation as the main growth promotion force in the twentieth century. Material accumulation then retains importance mainly as a vehicle for the accumulation of new knowledge. We can anticipate less and less reward for the brute act of saving and more and more for produced knowledge, innovation and education under what may be called the epoch of technocratic capitalism.

IV. *Relevance to the Less Developed World*

What light, if any, does this account of historical events in Western Europe shed on the problems of growth in the developing world? While as Kuznets has pointed out,[17] there are striking cultural and economic differences between the now developed Western countries and the contemporary developing non-Western world, the answer lies, obviously, in the feasibility of telescoping Western European experience. Several factors may be mentioned in this respect.

Parallel Development. In historical perspective, the Western experience with modern growth is based on the transition from mercantile agrarianism to industrial (and ultimately technocractic) capitalism. For the contemporary developing countries, modernization follows a different epochal sequence. We must remember, for example, that the period of mercantile agrarianism in Western Europe also coincides with the beginnings of the colonial period in the overseas territories and that, in fact, colonialism can be viewed as the foreign graft on what was essentially European development. Beginning with the end of World War II these very countries then faced the problem of achieving modern growth from the foundation of a colonial heritage under open agrarianism. We need not go into the details of that heritage comprising the familiar pattern of "enclave growth" under colonialism; i.e., (1) a static dualism as between traditional agriculture and the commercialized enclaves; (2) the export orientation of that enclave controlled mainly by entrepreneurs from abroad;[18] (3) the compartmentalization of growth and the generally stagnant character of the rural backyard. It is sufficient to point out that transition into modern growth from such a background added up to a considerable handicap in the light of Western experience.

The most important handicap is that the social overhead capital formation—i.e., the tree-star structure—which had taken place in Western Europe, gradually and unobtrusively over several centuries, had not really come into being in much of the less developed world by the end of colonialism. The construction of the feeder roads, highways, and communication systems (the tree structure) and the rural community centers and urban amenities (the star system), spreading over nearly three centuries (1500–1750) all of which could and did make such a substantial difference in facilitating the transition of Western Europe, did not take place in the overseas societies which essentially retained their position as agrarian appendages till World War II. This is not to say that substantial overhead capital formation in the overseas colonies did not take place. But it is also true that such investments, largely directed to facilitating the procurement of cheap labor for the enclave, and the outward flow of minerals and raw materials from it had relatively little impact on the bulk of the domestic agrarian economy.[19] In general, colonialism can be said to have intervened and thus prevented the occurrence of a parallel transition to industrial capitalism.

---

[17] S. Kuznets, "Developed and Underdeveloped Countries: Some Problems of Comparative Analysis," *Zeitschrift für die gesamte Staatswissenschaft,* 1968, 124(1).

[18] As J. S. Mill (*Principles of Political Economy,* London, 1929, pp. 685–86) put it, these enclaves were viewed as places "where England finds it convenient to carry on the production . . . of a few tropical commodities" and not as "countries with a productive capital of their own."

The first practical lesson to be learned then is that the difficulties encountered by contemporary underdeveloped countries are related to their inheritance of a weak rural development base. This weakness is measured partly in terms of the deficiency of physical overhead capital and partly in terms of the weakness of traditional rural organization.

The lesson which we have learned from Western history is, however, not entirely negative. Our tree-star structure emphasis, if accepted, suggests that the focal point of a policy of modernization of agriculture should be much more centered on the general notion of contact or "connectedness" which has internal as well as external dimensions. Externally, the agriculture sector must physically be in touch with the relatively modernized industrial sector. Internally, farmers must learn to have more contact with each other at the local community level. Application of this principle implies that, *ceteris paribus*, the modernization of agriculture may be facilitated where the population is homogeneous and fairly concentrated across land space as well as the beneficiaries of centuries of prior agrarian civilization (e.g., Japan, Taiwan, Korea). For a thinly populated large country with a weak agrarian cultural heritage, modernization of agriculture (and hence of the economy as a whole) is a much more difficult process—which can, however, be hastened to the extent that the problem can be solved by more "contact" and the creation of the requisite infrastructure.

The Time Dimension. In attempting to build up the necessary tree-star structure, a contemporary underdeveloped country is faced with the additional handicap of having to telescope centuries of Western European experience into a few decades. Moreover, that Western experience gave us no clue as to whether, or how, the improvement in organizational efficiency and/or the augmentation of rural overhead capital can be accomplished in a short time. On this matter of shortening the time span for rural modernization the developing country is faced with a new task—never before attempted with the possible exception of nineteenth century Japan.

It is indeed doubtful that the modernization of traditional agriculture can be fully telescoped to the extent that such a cultural accomplishment can be gauged in terms of a few five-year plans. The content of the tree-star structure strongly suggests that rapid rural development requires *inter alia* a mode of organization involving a peculiar combination of both individualism (i.e., individual initiative based on market discipline) and collectivism (i.e., collaborative effort at the community and grass-roots level). It is extremely doubtful that the strengthening of these weak links, e.g., the fostering of

---

[19] For a fuller description of the typical colonial mechanism at work, see the authors' "Agriculture in the Open Economy," paper presented to the Universities-National Bureau of Economic Research Conference on the Role of Agriculture in Economic Development, Princeton, N.J. (to be published).

"collectivism" through community development and agricultural extension services, is a matter that is responsive to short-run policy stimulation at the central government level.

The impotence of central government bureaucrats in dealing with rural programs is, itself, a matter which is best understood in historical perspective. Given their colonial heritage, the "modern" government machinery in developing countries may be irrelevant because its administrative experience is limited to the urban enclave within the compartmentalized dualistic system of that era. After decolonization, while the civil servants in the "British tradition" are fairly capable of handling urban related household and commercial activities (law, bookkeeping, licensing, banking, warehousing, etc.), they are inexperienced in industrial capital activities and both inexperienced and disinterested in rural administration. Add to this the customary identification of rural households of the central government with colonial tax collection, and the difficulties of generating the basis for a mutually reinforcing agro-industrial growth process will be realized.

Many of the contemporary underdeveloped countries will probably have to look to the history of Japan, both at home and as a colonizer, in search for clues as to the proper role of a central government in rural development. At home the reliance on local community action in raising resources for overheads and services to be used at the local level is instructive. Abroad, the success story of the agricultural revolution in Taiwan (a former Japanese colony) provides a unique experience to others. By general consensus, the farmers' organizations—first established as "top-down" instruments but later evolving into a unique combination of individual initiative and cooperative community effort—represented the single most important factor contributing to that success.These farmers' cooperatives, a cultural heritage of Japanese colonization, have a history of at least forty or fifty years. In telescoping Western experience, an important lesson of history is that agricultural modernization can be a time-consuming process even under the best of circumstances.

V. *Conclusion*

In long-run historical perspective, development in the second half of the twentieth century will probably be remembered as a phase of transitional growth as the economy moves from the long epoch of colonialism to that of modern growth. The Japanese experience seems to indicate that this transition process itself may be expected to take us at least into the twenty-first century, perhaps even longer, if the developing countries are to make a successful transition to economic maturity. This transition process is shaped by two basic factors. Exogenous to the underdeveloped country is the availability of new technology, as continuously thrown up by progress under technocratic capitalism in the industrial countries of the West. Internally, the essential growth phenomena are learning processes with respect to both technological borrowing choices from abroad and with respect to the gradual augmentation and improvement of the resources base at home.

What we have witnessed over the past twenty years is the very beginnings of this process of attempted transition. In fact, much of what goes under the name of contemporary development economics represents an attempt to gain a better understanding of the rules governing that transition. With so much required for synchronized intersectoral growth, the enhanced understanding of an ideal sequential ordering of events which can be distilled from history could be very helpful.

For example, those miscellaneous policy-tinged topics such as foreign aid, inflation, stabilization, absorptive capacity, which have taken up so much of our attention since World War II can be seen increasingly to be of only transient interest. Other topics relating more directly to the interaction between the two sectors, e.g., intersectoral labor allocation and the reduction of labor inefficiency, the institutionalization of intersectoral financial intermediation, government fiscal and overhead creation capacity, on the other hand, are likely to loom ever larger with time. We have attempted to single out one especially crucial development issue; namely, that of the modernization of agriculture in a multisectoral growth context. Western European history has indicated that this is a long-term issue and must reasonably be solved "first," if successful transition to maturity is to be expected.

Moreover, our realization of the importance of a minimum tree-star structure for mutually reinforcing growth is based on the recognition that intersectoral connectedness is essential for the generation and efficient allocation of an adequate agricultural surplus at each point in time. Once a minimum tree-star structure exists, in addition to the flow of commodities, the flow of saving through relatively "near" or familiar financial intermediaries, and the flow of human resources in search of higher productivity employment, can take place.

The contemporary LDC, of course, faces a very difficult task in its attempt to create the necessary structure overnight. Public sector action seems to be indicated, but not only resources but also the ability of civil servants to make the "right" decisions, in a hurry, is limited. Developing countries cannot wait for the gradual evolution from feudal to individualistic and ultimately individualist/collectivist mixes in the way resources are organized. Instead, the attempt to quickly restructure the postcolonial economy has led many of them into a rather frantic type of public sector interventionism. Only in very recent years is the bankruptcy of that policy becoming evident and have the lessons of history been taken to heart. While the modernization impulse may well have to come from outside agriculture, without the mobilization of that sector and its full interaction with the rest of the economy via an ever broadening net of human and market participation, development is difficult to sustain. Nor are the foreign resources needed to continue the costly "big push" industrialization policies of the past likely to be forthcoming.

For future generations the contemporary experience of developing countries will prove to be unique in at least one essential respect; namely, that the modernization of traditional agriculture is inseparable from the domestic industrialization process or from the importation of modern technology from abroad and its assimilation. This historical vision points to the need of a new emphasis in research; namely, the necessity of an intersectoral approach—which is illustrated by the "contact" approach to agricultural modernization. What we need to work towards is a more precise anatomy of the transition to economic maturity—possibly differing by type of society[20]—in terms of a sufficient explanation of an ideal sequential order of events—as well as in the logical necessity of that sequential order.

---

[20] Space constraints do not permit us to say more about the "typological" lessons of history.

THE CONTENT OF DEVELOPMENT ECONOMICS

By LLOYD G. REYNOLDS
Yale University

Economic development is an activity in search of an intellectual framework. The impetus to development studies was the appearance after 1945 of many new nation-states, with low income levels and primitive economic systems, able for the first time to assert their needs in international forums; and the political and commercial rivalry among the richer nations, which inclined them to make some positive response to these needs.

But a domain of practical activity does not constitute a scientific study; and the basis for this new subject remains unclear. Opinions range all the way from those who take the existence of development economics for granted to those who deny that there is any such thing.

Let me stake out a centrist position in this spectrum. Those of us who work on the less developed economies cannot yet claim the coherence of subject matter, accepted analytical apparatus, and wealth of empirical observations which characterise the older economic specialties. Courses and texts in economic development do not show the homogeneity of courses in international economics or public finance. There is, however, both the potentiality and the emerging outline of a true specialty, which will assume an increasingly solid place in the curriculum. It is not a passing fad which will go away if we wait a decade or two. It will not go away because the less developed economies will not go away, nor will their problems become less interesting and important.

This emerging specialty will not, however, be symmetrical with the older functional or sectoral specialties. Its domain is a certain range of national economies in certain regions of the world. Work on these economies requires mastery of existing macro- and microeconomic tools, together with new tools which remain to be invented. In this respect development economics resembles the study of socialist economies, or the study of economic history, rather than a specialty confined to one sector of the economy. It interpenetrates with the functional specialties rather than standing alongside them.

Two Preliminary Comments

Having asserted that economic development is at least a potential subject, I am duty bound to say what this subject is about. But before this, let me make two ground-clearing remarks.

One can argue that analysis of the LDC's consists mainly of applying standard economic tools in familiar ways. I understand that one leading graduate school has implemented this view by abolishing separate courses in economic development and substituting a "development section" in the standard courses on public finance, labor economics, industrial organization, and so on.

There is an element of truth in this argument. In shifting from the American economy to that of Chile or Pakistan we do not throw away our present tool kit. Nor can one dissent from the view that the standard functional courses should have a larger non-Western content. Such courses can gain in generality and power by being stretched to cover the LDC's and the socialist countries. But I do not think this is all that needs to be done. There is a place for special courses on socialist economies, which view them in the round and explore such distinctive features as the planning mechanism. There is room for similar specialized treatment of the less developed countries, along lines which I hope to indicate.

While familiar theoretical concepts provide a starting point for analyzing a less developed economy, it is only a starting point. The institutional and behavioral peculiarities of the LDC's require a significant amount of new tool construction. For example, Wallich, Seers, and others have shown that economic fluctuations in the LDC's are usually induced by swings in export receipts rather than in domestic investment; and that an understanding of their impact on the economy requires microanalysis of the main export industries. Again, Brazilian inflation really is different from American inflation, analytically as well as empirically. Peasant households cannot be presumed to behave like midwestern farm households. A student of the less developed economies has to be partly a tool-builder, or tool-adapter. People who have worked on this kind of economy for a long time are best prepared to refashion theory in useful directions.

It is also a misconception to think that one can handle the LDC's by tucking them into a general course on growth theory. Early economic growth in poor countries is a sufficiently different phenomenon from growth in mature industrial economies that it is misleading to package them under the same label. The standard assumptions of modern growth theory—one or two products, free mobility of factors, competitive pricing, constant returns to scale, well-behaved production functions, a closed economy—are quite fanciful even for the developed countries. To extend such assumptions to the LDC's is not at all plausible.

Moreover, from a LDC standpoint, these models do not pose the right questions. They take most of the interesting variables as given. The model then determines the rate of steady-state growth and whether, if the economy is off its equilibrium path, it will or will not converge toward it. Comparative dynamics is the name of the game. In a LDC, however, the interesting questions are what determines the rate of increase in factor supplies, the speed and direction of technical progress, and the increase of total factor productivity. The givens of the standard growth problem must be investigated as variables.

My second preliminary comment is that there is considerable confusion over whether the center of gravity of development economics lies in positive economics or in normative economics. Much of the literature embodies a "how-to-do-it" approach to the subject; and some writers have even professed to know the answers. The secret of growth has been viewed as lying in infusions of physical capital, in human capital formation, in institutional change, or in still other directions.

Now certainly some economists should be working full time on immediate policy problems, and some useful normative concepts are available. Cost-benefit analysis is perhaps the most useful single tool. Input-output analysis can be used to explore consistency of projections for particular sectors of the economy—for example, modern factory industry or the educational system. Long-term budgeting and programming techniques within the public sector are obviously useful.

But surely our central task is to understand the economic mechanism of the less developed countries. We are scientists before we are engineers. Our comparative advantage lies in building a systematic body of knowledge about how an economy operates, which improves our power to predict the consequences of policy actions.

Practical problems have always been important in stimulating new developments in economic thought. But between initial problems and policy conclusions there has typically intervened a gradual and arduous development of positive economics: definition and measurement of important variables, analytical models intended to replicate important aspects of behavior, framing and testing of specific hypotheses, gradual emergence of a picture of the economy in action. I do not see why this should be any less true in the study of less developed economies. As Keynes remainded us, "Economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind. . . ." I predict that the maturing of development economics will mean increased concentration on building such an apparatus for the LDC's; and that as they do this, development economists will win greater practical usefulness as well as intellectual respect.

I pass now to the constructive part of my task. Viewing development economics as a branch of positive economics, what is it really about? What kinds of work on the LDC's may yield a better understanding of their operation and lay a firmer basis for policy making? Let me suggest briefly four useful lines of activity: construction of growth models adapted to the less developed countries; empirical analysis of early economic growth in particular countries; comparative cross-sectional studies; and microanalysis of economic behavior.

Models of Early Economic Growth

I have already suggested that most of recent growth theory has little relevance to the LDC's, and that a different species of theory is required. In my judgment, the most useful "LDC growth model" produced to date is the Fei-Ranis model. Its authors would be the first to agree, however, that we stand only at the beginning of this kind of activity. Looking down the path which they have helped to open, I visualize several lines of development.

First, we all realize that the universe of LDC's is very heterogeneous. There are numerous "types" of LDC, though no one can yet say confidently what we mean by a "type." One obvious basis of classification is by relative factor availabilities. Many LDC's are "labor surplus economies," but others are not. The numerous countries which still have open frontiers might be termed "land surplus economies." There are even a few oil-rich countries which can be considered "capital-surplus economies." Size and openness is another important dimension, along which one can array LDC's from Jamaica to India. Seers has suggested an output-mix classification, based partly on the importance of industry relative to primary production, partly on the diversification of output within each of these categories. In any event, we should not expect any one growth model to have equal explanatory power for every kind of LDC. Rather, we should look toward a "family" of models tailored to particular circumstances.

Next, we need to experiment with a finer subdivision of sectors. Two sectors are certainly better than one, but four may be better than two. I have suggested elsewhere[1] that it would be useful to distinguish the public sector, "modern" private industry, traditional urban activities (trade, services, handicrafts), and agriculture. My mathematician friends tell me that four-sector models are extremely difficult and may not be susceptible of general solution. But it may still be useful to tinker with models which come closer to reality, even if we can only develop illustrative special cases or run simulation experiments.

Third, trade and capital movements are central facts of life in most of the LDC's.

---

[1] "Development with Surplus Labor: Some Complications," *Oxford Econ. Papers*, Mar., 1969.

There is general recognition that we really need open-economy models, which portray the interaction of growth and trade. Here the existence of a long tradition of trade theory is an advantage. On the other hand the extreme simplification of most trade theory, leading to limited ability to predict trade flows, is a considerable weakness. Work in this area can perhaps contribute as much to trade theory as to building better models of economic growth.

Finally, let me note that virtually all growth models—developed or underdeveloped—assume that growth is already under way "before the curtain rises." They do not tackle the intriguing question of how growth gets started in a previously stagnant economy. This problem obviously involves noneconomic variables, some of which are difficult to quantify; and so economists tend to hold back from it. But there is little indication that political scientists, social anthropologists, or others are going to do the job for us. So I believe we should be venturesome enough to make some forays into this borderline area. Let's have a little "economists' imperialism." The work of Everett Hagen, Irma Adelman and Cynthia Taft Morris, and others indicates that some economists are already moving in this direction.

The Movement of Economies Through Time

A second task is analysis of how particular economies have grown (or failed to grow) over time. There are forty to fifty LDC's of some size and consequence. Each of these provides a laboratory in which one can test models of early economic growth.

The Economic Growth Center at Yale initiated a major program in this area in 1961, which is now (hopefully) approaching completion. We selected twenty-five LDC's, reasonably well distributed by continents and including most of the larger countries. Within the foreseeable future, we shall have a shelf of books analyzing the growth experience of these countries, with particular attention to the post-1945 period but going back in some cases to 1900 or beyond.

This kind of work is not easy to categorize. One might call it economic history. More precisely, one might call it quantitative, analytical, Kuznets-type economic history. We have emphasized analysis of quantitative information about these economies; and each of the published studies will contain a substantial, annotated data appendix.

Given adequate information, what would one like to find out? Agriculture usually bulks largest in the economy. So one needs to analyze agricultural output and the distribution of output among crops and among major uses. To the extent that input data are available, one can explore how far output expansion is due to increased factor use and how far to increased factor productivity. Where factor productivity has risen, how and why did it rise? On the income side, one would like to know how income arising in agriculture is distributed between landowners and cultivators, where these do not coincide, and how much is reinvested in agriculture or other sectors.

Just as agriculture is the largest productive sector, labor is the largest input. So one has to analyze the labor force and wrestle with concepts and measures of "unemployment" and "underemployment." The test of successful development is an employment test as well as an output test. If we cannot say whether the ratio of employed labor time to available labor time is rising or falling, we cannot tell whether the economy is meeting the employment test. Nor can we tell what is happening to man-hour output and other productivity indicators.

Economic growth and structural change imply reallocation of the labor force and a rise in labor force quality: skill level, educational level, motivation, and personal efficiency. Study of the quality dimension of the labor force leads into analysis of the educational system, viewed as the dominant influence on the supply side of the labor market. In the early stages of growth, something called "the skill constraint," or "limited absorptive capacity," is said to limit the growth rate. What is the nature of this constraint, and how is it relaxed (or not relaxed) over the course of time?

The importance of the international sector is obvious. Export performance and the reasons for it need to be analyzed. Imports are usually controlled by an elaborate structure of exchange rationing, import quotas, and tariffs, not necessarily well rationalized or internally consistent. The way in which this structure has affected domestic resource allocation is an important subject for study. Capital movements, private and public, gross and net, need to be analyzed in terms of their productivity contribution and the current and future costs which they impose on the economy.

The public sector is still small in most LDC's, but it contains many of the levers which (in principle) a growth-minded leadership can manipulate to move the economy forward. So one should examine the growth of the main types of public expenditure. Development strategy is supposed to involve a rise in the proportion of revenue devoted to fixed investment and other developmental purposes. In some countries this has happened. In others, including some with comprehensive development plans and many statements of good intention, it has not happened. We need to know why. The tax system has to be analyzed with respect to incentive effects as well as revenue-raising power. Infrastructure industries, which are usually organized as public corporations, present important issues of investment, output, and pricing policy.

Manufacturing is also a growth sector whose expansion normally outpaces that of national output. The issues in this sector have been discussed extensively in the development literature and little need be added. We need obviously to know much more about such matters as the sequence in which new industries normally appear in a LDC; how far this sequence can be altered by policy, and the gains and costs of so doing; the borrowing and adaptation of technology; the sources of capital and entrepreneurship; the disposition of industrial profits; the rate of growth of employment, and the interrelations among employment, productivity, and wage policy.

A picture of changes in particular sectors does not necessarily add up to an understanding of the economy. There is interaction among sectors even in a stationary economy, and in a growing economy one would expect such interaction to be increasing. This interaction will show up in intersectoral flows of commodities, labor, and finance. So mapping these flows is an essential step in analyzing the economy.

Having done this, and assuming that growth has been occurring, one can proceed to interpretation and explanation. In the back of one's mind will naturally be hypotheses and models of how early economic growth may occur. But these are still sketchy, as indicated earlier, and if historical study had to wait for their perfection it would wait a long time. In this area, as in other branches of economics, progress will doubtless come through an interaction of theorizing and investigation, with crude hypotheses serving as a guide to research, which in turn suggests new or refined hypotheses.

Comparative and Cross-sectional Studies

Let me pass to a different kind of investigation. Suppose we want to say something, not just about what happens in Brazil, but about what happens in the LDC's generally. Or suppose we have a strong interest in a particular set of economic relationships. Then we will probably turn to testing hypotheses by cross-country econometric analysis.

To take an illustration from the developed countries: the rate of increase in money wages can be explained reasonably well by such variables as the unemployment rate, the profit rate, changes in these two rates, and changes in the consumer price index. We have data on the relevant variables for most of the industrial countries. So it is possible to derive an "international Phillips curve" from cross-country data, and this has actually been done.

It is easy to think of interesting hypotheses which might be tested in the less developed countries. For example:

1. The hypothesis that production functions are identical among countries, which still underlies much of trade theory.
2. The hypothesis that property income is saved while labor income is not, and that this contributes to differences in national savings rates.
3. The hypothesis that a country's pattern of manufacturing output is explained mainly by its population and per capita income.
4. The hypothesis that industrial wage rates in the LDC's are pushed above the supply-price of labor by institutional pressures, that this leads employers to substitute capital and management for labor, and that this results in a slow growth of industrial employment relative to industrial output.
5. The hypothesis that new industries in the LDC's tend to imitate developed-country technology, which may be quite inappropriate to the factor endowment of these countries.
6. The hypothesis that output of traded goods relative to nontraded goods, while it may rise initially as per capita income rises, will eventually decline.

Much of the progress in development studies over the next decade or two will probably come from such cross-country analysis. But progress will not be easy. I will only mention such a crass matter as data difficulties, which are still the most serious obstacle. The social marginal product of people who devote their lives to data improvement is high, and it is too bad that they receive such a low status ranking in our profession.

Beyond this, regression analysis assumes that variables are being defined and measured uniformly, and that the relations among them are of the same character throughout the universe being studied. This is less plausible for intercountry than for intracountry analysis. I am skeptical of cross-section results which include countries with very different income levels and economic structure—say, the United States and the Republic of the Congo. Even the universe of less developed countries may be too heterogeneous to be treated as an entity, and we may need to analyze subgroups of countries.

Then there is the question how far one can use cross-section results to interpret movements over time. This is a familiar problem from U.S. investigations, but it is probably more serious in cross-country studies. One may well hesitate to take cross-section findings as definitive unless they are supported by evidence from national time series.

Let me note finally that there are useful kinds of comparative study which may not be susceptible to refined statistical techniques. Take the question whether there is a Latin-American "type" of economy. This may be an unanswerable, or even a foolish, question. But it may not be. This kind of question requires qualitative, institutional analysis as well as statistical comparisons.

Microanalysis of Economic Behavior

Let me turn finally to what will undoubtedly, in sheer volume, be the largest kind of research activity in the less developed countries. This is detailed analysis of limited problems in a particular sector of a particular economy. After all, economic knowledge in this country has not progressed by people setting out to investigate the American economy. Rather, they have set out to investigate price-determination in the aluminum industry, or the determinants of fixed capital investment in manufacturing, or the incidence of the corporate income tax. After thousands of such studies, we can begin to see the American economy in the round, in a way which we cannot yet do for Peru or Ghana.

Let me illustrate from agriculture, which is the largest industry in almost every LDC. The operation of the peasant household is central to an understanding of the economy. Several competing models appear in the literature: the "inert peasant," the "lazy (or satisficing) peasant," the "maximizing peasant." But we do not know which of these models is most plausible, nor can we find out without more empirical study.

The system of land ownership and the division of output between owner and cultivator may have important effects on labor input, choice of products and techniques, and receptiveness to technical change. Proposals for changes in the tenure system are warmly debated in many countries. In most cases little is known about the economic consequences of one system or another. Yet quantitative analysis is possible. One occasionally finds almost a laboratory situation, where the same crops are grown in the same area, under two or more tenure systems. In such cases input-output relations can be examined, and one can ask whether land tenure per se has effects which can be segregated from those of other variables.

There is a large literature, mostly of a speculative character, on the possible existence of "surplus labor," "redundant labor," or "disguised unemployment" in the agricultural sector. It is doubtful that further verbal battles on this front can yield any positive product. But there is a shortage of studies in which precisely framed hypotheses have been confronted with relevant data. Much of the verbal argument, indeed, relates to a situation which is rare in reality: that of a declining farm labor force. The common situation in the LDC's, however, is that high population growth is swelling the farm labor force. The interesting problem for study is how this growing labor force is absorbed (or not absorbed) into the rural economy and what happens in the process to labor inputs per acre and to production methods.

There is a growing body of evidence that, where alternative crops are feasible, peasant producers are responsive to changes in relative prices. But this may be merely a substitution among outputs with no increase in aggregate output. Much less is known about how aggregate output responds to increased income possibilities. To put the point differently: what proportion of a potential increase in output must be left with the cultivator to persuade him to produce the output? Some material incentive is required, but there is little evidence on how much.

We have not chosen agriculture for illustration because it has been especially underinvestigated. On the contrary, there has been more careful research on agriculture in the LDC's than on any other sector. Knowledge is slight only in relation to the size and complexity of the industry. There is also a substantial literature on foreign trade, capital movements, and related matters. On other sectors there have been few significant pieces of research. But this means that the marginal yield of research effort is still high; and if economists behave rationally, careful microanalysis in the LDC's should grow rapidly.

Among other things, there is here an almost limitless field for doctoral dissertations. There is no reason for a graduate student in public finance to pursue a minute and tedious study of some aspect of the Nevada sales tax when he could be studying how to extract revenue from farmers in Pakistan. The latter problem is more interesting scientifically, more important practically, and doesn't make him any less a public finance man.

So I return in conclusion to a point made at the outset: the study of economic development interpenetrates with the conventional specialties rather than standing in a watertight compartment alongside them. It is undesirable and unfeasible for development specialists to draw a jurisdictional fence around the LDC's and announce: "This is our territory. All others stay out." There is no reason why a labor economist shouldn't study wage structures and labor allocation in Chile or Zambia as well as in the United States or Sweden. Such an infusion of talent from the traditional specialties of economics will advance our understanding of the LDC's. It should also produce broader economists and more interesting courses throughout the economics curriculum.

This paper has been couched in research terms, because it is through the accumulation of research that development economics will take firmer shape as a branch of economic study. It has been couched in forward-looking terms, because the subject is still more potential than actual. I hope, however, to have given you reasons for concluding that it would not be wise to sell development economics short.

INSTRUMENTS AND GOALS IN ECONOMIC DEVELOPMENT

By Irma Adelman, *Northwestern University*
Marsha Geier, *Yale University, and* Cynthia Taft Morris, *American University*

I. *Introduction*

The present paper is designed to shed empirical light upon the nature of the simultaneous relationships among instruments and goals in the process of national development and modernization. For this purpose, it applies the technique of canonical analysis to data characterizing the social, political, and economic attributes of seventy-four non-Communist underdeveloped countries. The basic question which the analysis seeks to answer is: which national goals can be most reliably achieved with the set of policy instruments at the disposal of developing nations.

The significance of this question is highlighted by recent revisions in the goals of U.S. foreign economic assistance to include objectives which may be more competitive than complementary with the goal of economic growth, narrowly conceived. Specifically, in the Foreign Assistance Act of 1966 and again in the Acts of 1967 and 1968, the Agency for International Development was enjoined to take explicit account of the social and political implications of its economic assistance programs and to favor those countries and projects which generate greater economic and political participation by broad segments of the population. Thus the Agency must make choices among competing goals of foreign assistance and must assess and give weight to the social and political by-products of economic development policies.

Canonical analysis can aid the making of choices between development policies by indicating those goals which can be most reliably achieved with a given set of policy instruments. In the present paper, the canonical technique is employed to select from a set of the several goals of national policy in underdeveloped countries and a set of the policy instruments usually available in low-income nations those subsets of targets and instruments most closely associated with each other and to estimate the nature of the simultaneous relationships among the two subsets. While the findings of the current analysis can contribute to practical choices among goals and instruments, they must of course be supplemented by policy judgments which establish priorities among goals. This is because targets which are comparatively difficult to achieve may be extremely important and thus be worth pursuit in preference to more easily attainable objectives.

The results of the present investigation of goals and instruments in developing nations suggest rather unexpectedly that goals relating to a wider distribution of the benefits of economic growth are, on the average, quite reliably associated with the instruments of development policy available to the underdeveloped countries represented by our data. Indeed, the reductions in economic inequalities between various sectors and regions of underdeveloped countries produced by long-run structural transformations of the aconomy appear to be more closely associeted with instruments of development policy than is the goal of a more rapid rate of economic growth per se. In contrast, the wider distribution of opportunities for purely political participation appears, for the typical underdeveloped country in our sample, not very predictably related to available instruments of policy. It is only among nations which have achieved the highest of the three levels of socioeconomic development depicted by our data that greater popular participation in political life appears to be significantly related to avilable instruments of policy. Thus, our analysis suggests that, when the developmental goals of the average underdeveloped nation in our sample are ranked according to the dependability of the relationship between means and ends, a wider distribution of income ranks first, economic growth second, and greater political democracy third.

The present paper is divided into six sections. In Section II, we explain the methods used to construct the canonical model. Section III describes the nature of the data incorporated in the analysis. Section IV presents the results of the canonical analysis for the full sample of seventy-four countries and compares them with a set of conventionally derived structural equations. In Section V, we discuss a further set of canonical results for three subsamples representing successive levels of socioeconomic development. Finally, Section VI consists of the summary and conclusions.

II. *The Construction of the Model*

Canonical correlation is a technique of multivariate analysis developed by Harold Hotelling, which extends multiple regression analysis to accommodate a multidimensional dependent variable, thus permitting statistical investigation of phenomena which are too complex in nature to fit into the straightjacket of a single "dependent" characteristic. More specifically, canonical analysis determines simultaneously sets of weights for both a "dependent" and an "independent" battery of variables, such that the weighted sums correlate as closely as possible.[1] In other words, each weighted sum or "canonical variate" is a linear combination of the variables on its side of the equation, while each pair of variates is derived so as to maximize the correlation between the weighted sums. As many canonical equations involving pairs of variates can be derived as there are dependent variables, each successive pair being determined so that the intercorrelation between the variates in the pair is maximized subject to the condition that each new linear combination be uncorrelated with previous functions. In the event of only one dependent variable, the unique equation becomes the usual multiple regression function among standardized variables.

In most macroeconomic models, the relationship between the goals and the instruments of economic policy is analyzed by formulating a preference function for the decision-maker together with a structural model of the economy. Once the model has been estimated statistically, nontarget endogenous variables are eliminated by algebraic substitution until a set of reduced-form equations remains in which each target variable is expressed as a function solely of the policy instruments and of the other exogenous variables. The reduced-form equations are then solved simultaneously for those values of the instrument variables which are consistent with the desired values of the targets and with the given data.

In this paper we bypass the formulation of a structural model. Instead, we use a stepwise form of the canonical correlation procedure to estimate directly a hierarchy of reduced-form equations each of which relates a subset of the targets to a subset of the instruments. Since the successive equations in this hierarchy are extracted in descending order of the extent of correlation between the goal and instrument sets, they serve to indicate which targets and which instruments are most closely associated with each other.

---

[1] A mathematical derivation of the canonical model is given in Appendix A to this paper.

In the stepwise form of the canonical procedure developed for use in the present paper, we first specify the number of variables in the canonical functions. To select the particular variables which enter the equations, the matrix of intercorrelations between the full sets of potential instruments and possible targets is scanned by the computer for those pairs of targets and instruments with maximum simple intercorrelations. With these preselected subsets of dependent and independent variables we then construct a set of canonical functions. To test whether the inclusion of more instruments improves the canonical correlation, we derive augmented canonical relationships by adding to the starting set of independent variables, in turn, each instrument variable not already appearing in the canonical equations. Any variable which significantly augments the canonical correlation we then add to the basic set.[2] The attempts to enlarge the instrument set are continued until no further variable can be found which significantly increases the canonical correlation.

By augmenting the independent side of the canonical equation we also gain a feeling for the probable significance of the derived canonical weights. Only those variables with consistently large coefficients in all corresponding canonical relationships are treated as statistically significant.[3] This check on the reliability of the weights is important because no tests for statistical significance exist for the individual weights in a canonical function. As a further cross-check on the robustness of the results, the analyses were repeated with both $(n+1)$ and $(n-1)$ variables in the starting subsets of instruments and targets and the successive results compared for consistency with respect to the important variables and their respective weights.

III. *Data*

The data used for the current model are from a study by two of the present authors of the interaction between economic and noneconomic influences in the process of economic growth and development. These data reflect the economic, social, and political characteristics of seventy-four underdeveloped non-Communist nations over the period 1957–62, as defined by both statistical and qualitative information. Full descriptions of the definitional schemes together with the individual country classifications and the matrix of intercorrelations appear in *Society, Politics and Economic Development: A Quantitative Approach*.[4]

For the purpose of the present analysis, the variables are divided into two groups: goals and policy instruments. Broadly speaking, the target variables or goals have two characteristics: they reflect fundamental positive changes which increase the effective range of choice of the individual members of the population with respect to their economic, social, and political actions and they can only be measured and implemented indirectly. In contrast, the instrument variables are characterized chiefly by the fact that they are more or less directly manipulable by national governments and are generally regarded as desirable primarily because of the ends they serve. Of course, the division between goals and instruments is not completely clear cut. A characteristic which can be altered by state policy may be considered so generally desirable that it can become a goal in and of itself (e.g., the modernization of agriculture). In addition, the degree to which a characteristic is subject to direct manipulation by the government varies to some extent with the time span under consideration as well as with the vigor of the government's program.

---

[2] More specifically, we apply the significance test to the canonical correlation coefficient pertaining to the second canonical relationship and to the trace correlation coefficient. In our tests, the addition of a variable never augmented the significance of the canonical relationships by either test.

[3] Experimentation with successive runs suggests that, with samples and data such as those used in the present analysis, a coefficient of greater than .40 may generally be treated as significant for the first canonical relationship; for the second and third canonical functions coefficients of .45 or over may be treated as significant.

[4] Irma Adelman and Cynthia Taft Morris (Johns Hopkins Press, 1967).

The list of variables presented below is divided into two categories, with Y symbols for the goals and X symbols for the instruments.

Goals. Extent of social mobility (Y_1); extent of social tension (Y_2); strength of popular political participation (Y_3); degree of national integration and sense of national unity (Y_4); extent of political stability (Y_5); rate of growth of real per capita GNP (Y_6); size of the traditional agricultural sector (Y_7); level of modernization of techniques in agriculture (Y_8); extent of socioeconomic dualism (Y_9).

Instruments.[5] Extent of literacy (X_1); extent of mass communication (X_2); crude fertility rate (X_3); degree of administrative efficiency (X_4); extent of leadership commitment to economic development (X_5); extent of direct government economic activity (X_6); gross investment rate (X_7); level of adequacy of physical overhead capital (X_8); degree of improvement in physical overhead capital (X_9); degree of improvement in the tax system (X_{10}); degree of improvement in financial institutions (X_{11}); degree of improvement in agricultural productivity (X_{12}); change in degree of industrialization (X_{13}); extent of secondary and higher education (X_{14}); structure of foreign trade (X_{15}); character of agricultural organization (X_{16}).

IV. *Results of the Canonical Analysis for the Full Sample*

The results of the canonical analysis consist of four sets of equations: one set for the full sample of seventy-four underdeveloped countries, and one set each for three subsamples representing successive levels of socioeconomic development.[6] In this section we discuss the large sample canonical functions and compare them with a set of reduced-form equations derived from a structural model. We discuss the small sample canonical results in Section V.

It will be recalled that the starting subsets of variables from which the canonical functions are estimated are chosen by selecting those pairs of dependent and independent variables having the highest simple correlations, in descending order of the size of their correlation coefficients. These subsets are then augmented, in a stepwise fashion, as explained earlier. Since the augmented relationships never passed the significance tests in any of the applications reported upon in this paper, the starting and final subsets of variables turn out to be identical in practice.

For the full sample, the starting subset of policy instruments selected for the independent side of the equation consists of the adequacy of physical overhead capital (X_8), the rate of literacy (X_1), the extent of mass communication (X_2) the extent of secondary and higher education (X_{14}), the degree of improvement in financial institutions (X_{11}), and the character of agricultural organization (X_{16}). The initial subset of goals chosen for the dependent side comprises the level of modernization of techniques in agriculture (Y_8), the extent of social mobility (Y_1), the size of the traditional agricultural sector (Y_7), the extent of socioeconomic dualism (Y_9), the rate of growth of per capita GNP (Y_6), and the extent of popular political participation (Y_3).

---

[5] It should be noted that the list of instruments includes both measures of levels of long-term achievement (as of about 1961) and those of short-term improvements (1950–63). Indices of long-term attainment are classified as instruments only when major changes in a country's ranking can in practice be achieved in a relatively short period of time; such is the case, for example, with the rate of literacy (X_1) and the extent of transportation and power networks (X_8).

[6] For details on the make-up of these samples, see Adelman and Morris, *Society, Politics and Economic Development*, pp. 10–11, and pp. 167 ff.

First Canonical Equation. The first vector correlation equation relates a set of goals pertaining to a more equitable distribution of the material fruits of economic development to a set of instruments describing the extent of social overhead capital in both its physical and human aspects. (This equation has a multiple correlation coefficient of .97.) The canonical equation states that a dependent set of targets comprising primarily the level of agricultural productivity (Y_8), the extent of social mobility (Y_1), the size of the traditional agricultural sector (Y_7), and the extent of socioeconomic dualism (Y_9) is associated with an independent set of instruments composed mainly of the adequacy of physical overhead capital (X_8), and the rate of literacy (X_1): More specifically

$$
\begin{aligned}
(1)\quad & .43Y_8 + .63Y_1 - .47Y_7 + .43Y_9 \\
& + .10Y_6 - .01Y_3 \\
= {} & .73X_8 + .50X_1 - .06X_2 + .26X_{14} \\
& + .14X_{11} + .36X_{16} \qquad (\bar{R}^2 = .97)
\end{aligned}
$$

The goals which emerge as most closely associated with the instrument set thus depict broad structural transformations which result in the wider diffusion of the benefits of economic growth. The specific changes involved are: the growth of a modern sector and the subsequent spread of modern technology throughout the economy, summarized by the dualism indicator;[7] increased opportunity for individual participation in economic growth through greater social mobility; and increased participation by the agricultural sector in the material benefits of growth both through the commercialization of agriculture (which reduced the size of the subsistence agricultural sector) and through improvements in agricultural productivity.

The instruments most suitable for achieving the distributional goals of the first canonical equation are seen to be programs which expand the power and the transportation networks and the primary educational system of the society. These measures typically have widespread effects in improving the productivity of capital, labor, and natural resources and create significant external economies. They also portray primarily a widening rather than a deepening of social overhead capital because in our definitions of these indicators improvements in physical overhead capital and literacy consist mainly of the extension of facilities into rural and semirural areas rather than of the intensification of effort in the major urban centers.

Thus, broadly speaking, the first canonical equation represents the concomitant variation of goals promoting a more even distribution of the material rewards of economic development with a wider diffusion of facilities for improving the human and physical capital base of the economy.

The Second Canonical Equation. The second equation correlating canonical variates, which is statistically significant only at the 10 percent level, reflects a more subtle pattern of associations than does the first equation. The principal feature of the target side of the equation is a substitution between greater social mobility (Y_1) and a more rapid rate of growth of per capita GNP (Y_6). The instrument variables with the largest coefficients on the independent side are the rate of literacy (X_1) and the index of improvements in financial institutions (X_{11}). Specifically, the second canonical function is as follows:

---

[7] It should be noted that the indicator of the extent of dualism is scored in such a manner that, at the lower end of the scale, a positive change represents the increase in dualism which emerges as the expansion of a distinct modern sector imposes itself upon a predominantly agrarian society. At the upper end of the scale, a positive change with respect to the dualism indicator consists of reductions in the dichotomies between the traditional and the modern as elements of both eventually become intermingled throughout the economy and society. See Adelman and Morris, *Society, Politics and Economic Development*, pp. 23–25.

$$(2)\quad -.23Y_8 + .70Y_1 - .12Y_7 - .05Y_9 - .65Y_6 - .12Y_3 = -.16X_8 + .54X_1 - .06X_2 + .37X_{14} - .71X_{11} - .19X_{16} \qquad (\bar{R}^2 = .61)$$

The presence in this equation of a pattern of substitution between more rapid economic growth and greater openess of the economic growth process underlines the difficult choice which developing societies face between the achievement of maximum growth rates and the wider distribution of the benefits of economic progress. An examination of the relationship between the dependent and the independent sides of the equation suggests that the literacy variable contributes to the goal of social mobility, while improvements in financial intermediation promote the goal of more rapid economic growth.[8]

General Discussion. In summary, the picture presented by the full-sample canonical analysis is one in which the policy instruments at the disposal of the typical underdeveloped country represented by our data are most reliably associated with improvements in the distribution of economic welfare. The goal of raising the economic growth rate per se, although less strongly associated with available policy instruments, nevertheless emerges as predictably attainable at some sacrifice of social and economic equity. Finally, the goal of greater political participation does not show a significant dependable association with available policy instruments. Broadly speaking, therefore, our results suggest that the capacity of underdeveloped countries to induce higher growth rates and wider participation in the material benefits of growth by means of development policy is much more certain than is their ability to generate political change.

The finding that for the "average" underdeveloped country in our sample distributional goals are more reliably associated with the instruments available than is the goal of raising growth rates is rather surprising. One reason may be that, before a process of sustained economic growth can be initiated, basic structural transformations of the economy must occur which have as one of their principal by-products improvements in the distribution of income. The current canonical analysis suggests specifically that, until the expansion of the market and the spread of modern technology into the traditional agricultural sector have proceeded sufficiently far and until adequate economic opportunities for a class of indigenous entrepreneurs have emerged, government policy is unlikely to be able to significantly raise the rate of economic growth. Another reason for the present pattern of associations may be that, since sustained growth in low-income nations typically requires a detailed development design comprising carefully integrated elements, the uncoordinated concurrent application of a set of blunt policy instruments does not necessarily lead to significant increases in the rate of economic growth.

Another interesting feature of the full-sample results is the finding that the goal of increased popular political participation is not predictably related to available instruments. Specifically, our index of political participation appears with an important loading only in the fourth, statistically insignificant, canonical relationship. This result probably occurs because, in order for social and economic changes to have a significant impact in furthering the evolution of participant political institutions, a higher level of achievement with respect to such social characteristics as literacy, mass communication, and urbanization is necessary than has been attained by the average country in our full sample. This interpretation is consistent with our earlier finding that a significant relationship between indexes of socioeconomic change and political democracy appears only among underdeveloped countries at a relatively "high" level of socioeconomic development.[9]

---

[8] The literacy instrument has a positive sign because of its direct relationship to social mobility (which is a positive inclusion on the target side), while the financial indicator has a negative sign because financial improvements accelerate economic growth and the rate of change of per capita GNP is a negative inclusion on the dependent side. Thus, the nature of the policy mix on the instrument side of the second canonical relationship is determined by the direction and extent of the substitution between goals on the dependent side.

Comparison of Results with a Structurally Estimated Model. A comparison of the reduced-form canonical model of the relationship between instruments and goals with the results obtained from the application of a traditional econometric approach is instructive. For this purpose, a set of structural models was estimated to "explain" each of the four targets which appear with important weights in the first canonical equation presented above. The structural equations for each of these models were derived by means of stepwise regression. For each dependent variable, a set of potentially causal independent variables was first specified. The stepwise regression procedure was used to select from the causal set those variables which appear in each endogeneous equation of the model for the specified target variable.[10] Reduced-form equations "explaining" each of the targets were then obtained by solving simultaneously the relevant structural equations.

The reduced-form equation for the extent of socioeconomic dualism (Y_9) relates reductions in dualistic cleavages between modern and traditional sectors to major extensions of transportation and power networks (X_8), increases in literacy (X_1), and greater diversification of the structure of foreign trade (X_{16}):

$$Y_9 = 3.93 + .47X_8 + .27X_1 + .13X_{16} \tag{3}$$

The equation for the extent of social mobility (Y_1) associates greater opportunities for social and economic advancement with increased literacy (X_1), the weakening of political barriers to upward social movement (Z_2),[11] and the more rapid extension of finance to indigenous entrepreneurs (X_{11}):

$$Y_1 = 11.10 + .73X_1 - .08Z_2 + .113X_{11} \tag{4}$$

The third and fourth reduced-form equation, those for the level of modernization of techniques in agriculture (Y_8) and the size of the traditional subsistence agricultural sector (Y_7) relate major improvements in agricultural productivity and the commercialization of traditional agriculture, respectively, to a single set of influences: the expansion of transportation and power networks into the agricultural hinterland (X_8), the spread of literacy (X_1), greater diversification of opportunities for foreign trade (X_{16}), the wider provision of finance to agriculture (X_{11}) and the weakening of the political strength of the traditional elite (Z_2). Specifically,

---

[9] See Adelman and Morris, *Society, Politics and Economic Development*, pp. 257–63.

[10] A full description of the procedure used in deriving the model equations will be published in a forthcoming book by the present authors. The details of the structural equations composing the models for the current four target variables are given in Appendix B to this paper.

[11] The indicator of the political strength of the traditional elite (Z_2) is not included in the canonical analysis because it is not a clear-cut case of either a goal or an instrument; in the present derivation, it is regarded as predetermined.

$$Y_8 = 1.12 + .66X_8 + .28X_1 + .04X_{15} + .01X_{11} - .01Z_2 \tag{5}$$

$$Y_7 = 96.71 - .33X_1 - .43X_8 - .04X_{15} - .01X_{11} - .01Z_2 \tag{6}$$

Equations (3), (4), (5), and (6) provide us with reduced-form equations for each of the four important dependent variables which appear in the canonical model. In order to make possible a comparison of the two sets of results, the reduced-form equations estimated from the structural model are aggregated with the same weights derived for the respective dependent variables in the first canonical equation. The reduced-form equation aggregated with weights corresponding to the first canonical relationship is as follows:

$$\begin{aligned} &.43Y_8 + .63Y_1 - .47X_7 + .43Y_9 \\ &\quad = 54.61 + .71X_8 + .86X_1 \\ &\qquad + .09X_{15} + .01X_{11} - .07Z_2 \end{aligned} \tag{7}$$

It will be recalled that the first canonical relationship estimated was

$$\begin{aligned} &.43Y_8 + .63Y_1 - .47X_7 + .43Y_9 \\ &\qquad + .10Y_6 - .01Y_3 \\ &\quad = .73X_8 + .50X_1 + .06X_2 + .26X_{14} \\ &\qquad + .36X_{16} \end{aligned} \tag{8}$$

It is evident from a comparison of (7) with (8) that in both equations the level of physical overhead capital (X_8) and the extent of literacy (X_1) are the important instrumental variables explaining inter-country differences in the achievement of the specified goals. The coefficients of the instrumental variables in (7) and (8) are indeed remarkably close, especially when one remembers that the extent of secondary and higher education (X_{14}) and the level of literacy (X_1) are positively related, and that some of the weight attaching to X_1 in equation (7) is therefore captured by X_{14} in equation (8).[12]

Thus, the canonical technique used in this paper appears to offer a reasonably reliable method for estimating directly an aggregated reduced-form equation from a set of simultaneous-equation structural models. Furthermore, canonical analysis has an advantage over the usual econometric approach in that it can be used where too little information is available to specify the underlying structure on a priori grounds. Finally, the canonical method has the additional advantage that it provides information on the ranking of the reduced-form equations according to the dependability of the relationships between goals and instruments which the equations express.

V. *Results of the Canonical Analysis for the Small Samples*

The Low Sample. The first of the subsamples representing different levels of development for which canonical analyses were performed is composed, for the most part, of African nations south of the Sahara but also includes a small number of countries in Southeast Asia and the Far East.[13] These are countries at relatively low levels of socioeconomic development characterized by large traditional agrarian sectors and by formidable sociocultural barriers to economic growth.

The starting set of target variables for the low-sample canonical analysis chosen

---

[12] The regression equation relating the extent of secondary and higher education (X_{14}) to the literacy rate (X_1) is:

$$X_{14} = 12.62 + .79_1 \qquad R^2 = .64$$

The use of this equation to substitute for X_{14} in (8) raises the coefficient of X_1 in equation (8) to .69.

[13] The following are the countries included: Afghanistan, Cambodia, Cameroun, Chad, Dahomey, Ethiopia, Gabon, Guinea, Ivory Coast, Kenya, Laos, Liberia, Libya, Malagasy, Malawi, Morocco, Nepal, Niger, Nigeria, Senegal, Sierra Leone, Somali Republic, South Vietnam, Sudan, Tanganyika, Uganda, Yemen, and Zambia.

on the maximum correlation criterion comprises the extent of dualism (Y_9), the degree of national integration and sense of national unity (Y_4), the extent of social mobility (Y_1), the level of modernization of techniques in agriculture (Y_8), the rate of growth of real per capita GNP (Y_6), and the extent of political stability (Y_5). The instrument variables selected in the initial set are the adequacy of physical overhead capital (X_8), the extent of secondary and higher education (X_{14}), the degree of administrative efficiency (X_4), the extent of literacy (X_1), the gross investment rate (X_7), and the extent of leadership commitment to economic development (X_5).

The first of the three significant canonical functions for countries in the low group is primarily between the extent of socioeconomic dualism (Y_9) on the target side and, on the instrument side, the adequacy of physical overhead capital (X_8) and the degree of administrative efficiency (X_4); it has a canonical $\bar{R}^2$ of .80. Specifically,

$$\begin{aligned} &.87Y_9 + .08Y_4 + .24Y_1 + .21Y_8 \\ &\quad - .35Y_6 - .15Y_5 \\ &= .69X_8 + .13X_{14} + .69X_4 + .09X_1 \\ &\quad - .03X_7 + .14X_5 \qquad (\bar{R}^2 = .80) \end{aligned} \tag{9}$$

Thus, the broad view of the relationship between goals and instruments suggested by the first canonical function for the typical country in the low sample is that the growth of a distinct modern enclave and the associated spread of market-oriented activities into agriculture is the goal most predictably attainable with the means at the disposal of the national government. In the achievement of this objective, transportation facilities and government services are most dependably associated with successful performance.

The second canonical equation for nations in the low sample indicates the presence of significant substitution between the goals of socioeconomic and political integration (represented by Y_9 and Y_4) and the goals of more rapid economic growth (Y_6); parallel substitution is evident because two instruments for achieving integration (investment in physical overhead capital, X_8 and the development of secondary and higher education, X_{14}) and two instruments for accelerating economic growth (the size of the aggregate investment effort, X_7, and the degree of administrative efficiency, X_4). Specifically,

$$\begin{aligned} &.51Y_9 + .66Y_4 - .20Y_1 - .22Y_8 \\ &\quad - .45Y_6 - .06Y_5 \\ &= .51X_8 + .31X_{14} - .01X_1 - .64X_7 \\ &\quad - .13X_5 \qquad (\bar{R}^2 = .73) \end{aligned} \tag{10}$$

This pattern of associations is not surprising since it is well known that, in the short run, the difficult tasks of socioeconomic and political integration facing underdeveloped fragmented societies frequently conflict with the goal of rapidly raising GNP in the market sector of the economy. This is in part because, in low-income traditional societies, the major capital-widening investments which integrate the agriculture hinterland with the major urban center can often only be achieved by slowing down those capital-deepening investments inside the modern sector which typically weigh heavily in increases in measured GNP.

The final significant canonical equation for the low group depicts a positive relationship between the goal of political stability (Y_5) and the complementary instruments of expanding the secondary and higher educational system (X_{14}) and increasing the overall rate of investment (X_7):

$$\begin{aligned} &-.32Y_9 - .26Y_4 + .07Y_1 + .30Y_8 \\ &\quad - .19Y_6 - .83Y_5 \\ &= .16X_8 - .64X_{14} + .16X_4 + .44X_1 \\ &\quad - .51X_7 - .28X_5 \qquad (\bar{R}^2 = .48) \end{aligned} \tag{11}$$

Investment tends to promote political stability because, by raising economic growth rates, it enables the political system to better satisfy the rising popular aspirations toward greater material welfare. The wider acquisition of educational skills contributes to greater political stability by furthering the social and political integration necessary for the successful mobilization of widespread political support for the national government.

Viewed as a whole, the canonical results for countries in the low group indicate a hierarchy of goals, ranked by the reliability of their association with the policy instruments at the disposal of national governments. First is the growth of distinct modern sectors within predominantly traditional societies and the related wider spread of market-oriented enterprise. Second are concurrent advances with respect to the economic and the political integration of the agricultural hinterland with the modern urban sector; these are typically achieved, however, at the cost of somewhat lower rates of growth of measured per capita GNP than would prevail if development policy were concentrated in the modern sector alone. The third goal in the hierarchy is the maintenance of the stability of the political system, a prerequisite for modernization in low-income countries characterized by fragmentation of both political and economic processes.

In general, therefore, in countries at the lowest level of development, the establishment and diffusion of modern economic organization is more dependably related to available instruments than the acceleration of economic growth per se. The raising of economic growth rates is nevertheless predictably associated with available instruments, but is usually achieved at the cost of fewer resources devoted to integrating traditional agricultural society with the more modern urban complexes. As in the full-sample analysis, the goal of wider popular political participation does not show a significant dependable association with available policy instruments.

The Intermediate Sample. The intermediate sample is characterized by transitional societies in the throes of social, economic, and political modernization most of which have broken through the major socioeconomic barriers to modernization present in the typical country in the low sample. The intermediate group includes, among others, countries from the Middle East and Mediterranean, South Asia, the Far East, and Latin America.[14]

The initial sets of six targets and six instruments selected for the intermediate group by the maximum correlation criterion are composed as follows: The targets are the extent of political stability (Y_5), the level of modernization of techniques in agriculture (Y_8), the extent of dualism (Y_9), the extent of social mobility (Y_1), the rate of growth of real per capita GNP (Y_6), and the size of the traditional agricultural sector (Y_7); the instruments are the extent of leadership commitment to economic development (X_5), the level of adequacy of physical overhead capital (X_8), the degree of improvement in agricultural productivity (X_{13}), the extent of secondary and higher education (X_{14}), the degree of improvement in physical overhead capital (X_9), and the character of agricultural organization (X_{16}).

Three of the canonical functions derived for the intermediate sample are statistically significant. The first one indicates the presence of significant substitution between increases in political stability (Y_5) on the one hand, and greater social mobility (Y_1) and reductions in the size of the subsistence agricultural sector (Y_7), on the other.[15] A parallel substitution appears between the instruments of increased leadership commitment to economic development (X_5) and a higher rate of improvements in agricultural productivity (X_{12}). Specifically,

---

[14] The countries are Algeria, Tunisia, Iran, Iraq, Syria, Jordan, Ceylon, India, Pakistan, Burma, Thailand, Indonesia, The Philippines, Bolivia, Guatemala, Ecuador, Honduras, Surinam, Ghana, Rhodesia, and South Africa.

$$
\begin{aligned}
(12) \qquad & -.61Y_5 - .28Y_8 + .31Y_9 + .52Y_1 \\
& \quad + .03Y_6 - .43Y_7 \\
& = -.80X_5 + .18X_8 + .43X_{12} \\
& \quad + .15X_{14} - .09X_9 + .33X_{16} \\
& \qquad (\bar{R}^2 = .86)
\end{aligned}
$$

This function suggests for countries at the intermediate level the foremost importance of their search for political stability, at the cost, however, of a slowdown in their rate of transformation from peasant societies into industrial nations (represented by the association between increased political stability and negative changes in the importance of traditional agriculture). The conflict between greater political stability and more rapid economic transformation of subsistence agriculture arises probably because the political coalitions of traditional land-based oligarchies, the military, and government bureaucracies frequently found in these countries tend to disintegrate when modernizing groups in the coalition institute land reform or other measures for improving agricultural productivity which distribute the benefits of economic change to the peasant rather than the landowner.

The inclusion in this equation of reductions in social mobility as a goal and increased leadership commitment to development as an instrument suggests further that the development process in these countries is usually carried out by oligarchic modernizing elites which seek greater political stability within a less open society. Such a pattern of leadership behavior is understandable for countries in which typically the primary obstacles to both economic and political development are serious political instability associated with extreme social tension and in which the potential political power base is rather narrow.

The second canonical function for the intermediate group emphasizes the goal of increased political stability (Y_5) together with the goal of expanding the modern economy and diffusing the growth process throughout the country (Y_9). The chief instrument for attaining these goals, according to this equation, is increased leadership commitment to economic development (X_5). Specifically,

$$
\begin{aligned}
(13) \qquad & -.78Y_5 + .20Y_8 - .45Y_9 + .19Y_1 \\
& \quad - .17Y_6 + .30Y_7 \\
& = -.78X_5 + .10X_8 - .19X_{12} \\
& \quad - .40X_{14} - .28X_9 - .33X_{16} \\
& \qquad (\bar{R}^2 = .75)
\end{aligned}
$$

Thus, the secondary intermediate relationship suggests that, within the framework of the basic choice by a country's leaders of a certain degree of political stability and a given extent of economic transformation of peasant society, economic modernization in the form of the expansion of the modern industrial sector together with the wider spread of modern technology into the traditional parts of the economy tends to increase political stability. The first and second canonical functions together express the rather complicated fact that, while the average country in the intermediate sample must, to a certain extent, choose between the pace of basic economic and social transformations and increased political stability, there is nevertheless an underlying long-run complementarity between the creation of a stable political system and the process of economic modernization. In the short run political stability and economic modernization compete for resources; in the long run, however, if agriculture is modernized, greater political stability results since economic aspirations in the countryside are more rapidly satisfied.

---

[15] It should be noted that the size of the traditional agricultural sector is scored so that negative changes consist of reductions in the size of traditional agriculture which in this equation are associated with decreases in political stability.

The third and final significant canonical function for the intermediate level suggests a certain degree of substitution between the goal of substantial increases in the extent of technological modernization in the agricultural sector (Y_8), and the goal of a more rapid rate of growth of per capita GNP (Y_6). If we look at the dependent–independent relationships, we find more rapid growth of average GNP associated primarily with an acceleration in the build-up of physical overhead capital (X_9), while technological modernization in agriculture is mainly related to the spread of secondary and higher education (X_{14}):

$$\begin{aligned} & -.38Y_5 - .70Y_8 + .37Y_9 + .09Y_1 \\ & \quad + .46Y_6 - .09Y_7 \\ & = -.09X_5 - .26X_8 + .26X_{12} \\ & \quad - .44X_{14} + .81X_9 + .01X_{16} \\ & \qquad (\bar{R}^2 = .57) \end{aligned} \tag{14}$$

There appears, therefore, to be a trade-off in these countries between economic growth conceived narrowly as changes in per capita GNP and the wider geographical distribution of economic benefits represented by the spread of modern technology throughout the traditional agricultural sector.

Taking the intermediate results as a whole, we find a ranking of national goals. Primary among them are the achievement of a stable polity and the transformation of peasant societies into industrial nations; each of these goals, however, is typically achieved at the expense of some slowdown in progress toward the attainment of the other one. The underlying policy emphasis on political stability is not surprising because, in general, countries in this group face significantly greater political and social instability than is usually present in the countries in the low and high samples. Next in the hierarchy of goals is the direct enlargement of the base of technologically modern and market-oriented economic activity. Finally come the acceleration of the growth of per capita income and the diffusion of technological progress within agriculture. Here again, however, there is a trade-off between two goals so that higher economic growth rates are commonly achieved only at the cost of some sacrifice in distributional equity.

Broadly speaking, the substitution relationships evident in the intermediate results suggest that many different development and modernization paths are available to the average country in the intermediate group. Their political leaders face choices both between the political end of a stable political system and basic economic transformations and also between the economic goals of more rapid growth of average income and a wider distribution of the benefits of growth. The above canonical analysis helps explain the diversity of patterns of socioeconomic and political development in intermediate-level countries found in factor analytic investigations by the present authors.[16] Given the varied paths of change suggested by the substitution relationships in the canonical results, and the fact that raising purely economic growth rates does not emerge as the primary target associated with other features of economic and noneconomic change, it is indeed not surprising that intercountry differences in rates of growth of per capita GNP were only weakly explained by the factor analyses for countries in the intermediate group. The canonical equations indicate that it is the presence of alternative political and distributional targets which accounts for the earlier factor analytic finding.

---

[16] *Society, Politics and Economic Development*, Chap. 6.

The High Sample. The high sample of countries includes sixteen countries from Latin America, six from the Mediterranean area, and three from the Far East.[17] On the average, they are significantly more advanced than the countries in the intermediate group with respect to social and economic development, although their progress toward economic maturity is as yet far from complete.

The starting subset of target variables for the high group is composed of the level of modernization of techniques in agriculture (Y_8), the rate of growth of real per capita GNP (Y_6), the extent of dualism (Y_9), the extent of political stability (Y_5), the extent of social mobility (Y_1), and the extent of political participation (Y_3). The starting instrument set includes the adequacy of physical overhead capital (X_8), the degree of improvement in financial institutions (X_{11}), the crude fertility rate (X_3), the degree of administrative efficiency (X_4), the character of agricultural organization (X_{16}), and the extent of mass communication (X_2). The canonical analysis performed with these variables yields three significant equations, the details of which are discussed below.

The first canonical relationship for the high group is between the complementary goals of spreading economic growth (Y_9) and increasing the extent of political participation (Y_3) throughout the nation, on the one hand, and reductions in the crude fertility rate (X_3) and changes in agrarian structure (X_{16}), on the other.[18]

$$\begin{aligned}(15)\quad & .40Y_8 + .07Y_6 + .71Y_9 + .34Y_5 \\ & + .06Y_1 + .45Y_3 \\ & = .53X_8 + .03X_{11} - .65X_3 + .25X_4 \\ & + .47X_{16} - .08X_2 \qquad (\bar{R}^2 = .92)\end{aligned}$$

These associations suggest that for countries which have achieved the level of social and economic development of most of the countries in the high group, the wider distribution of opportunities for participation in both modern economic activities and political processes is a primary goal attainable with the policy instruments typically available to government. The instruments for achieving these goals are both economic and social in nature. Changes in agrarian structure are undoubtedly included in the equation because institutional improvements in agriculture consist primarily of reforms of land tenure which increase incentives to engage in commercial agriculture by linking economic rewards with productive efforts. The other instrument variable with important weight is the crude fertility rate. The reason for its inclusion is probably that this variable is a proxy for the underlying changes in values and styles of life associated with reductions in fertility. Such changes work in the direction of greater individual participation in modern economic and political processes and are not otherwise captured by our indicators. The essence of these attitudinal transformations is that they lead to a widening of the sphere of life in which individual rationality and choice are applied to everyday decisions.

The second canonical function for the high group suggests the presence of a significant degree of substitution between the goals of political stability (Y_5) combined with greater social mobility (Y_1) and the goal of expanding the geographical area over which economic growth takes place (represented by the dualism indicator, Y_9); the independent side of the equation expresses a parallel substitution between the instruments of increased administrative efficiency (X_4) and major build-ups in physical overhead capital (X_8).

---

[17] The total twenty-five countries are as follows: Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, El Salvador, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Trinidad, Uruguay, Venezuela, Greece, Turkey, Cyprus, Lebanon, Israel, United Arab Republic, Japan, South Korea, and Taiwan.

[18] We have somewhat less confidence in the canonical relationships for the high group because they do not always maintain their principal characteristics in the seven-variable results.

$$\begin{aligned} & -.43Y_8 + .32Y_6 - .45Y_9 + .48Y_5 \\ & \quad + .58Y_1 - .02Y_3 \\ & = -.66X_8 + .08X_{11} + .33X_3 \\ & \quad + .54X_4 + .34X_{16} + .22X_2 \\ & (\bar{R}^2 = .85) \end{aligned} \tag{16}$$

This function suggests that for given levels of achievement in broadening economic participation, further geographic expansion of the impact of economic growth can only be attained at the cost of increased political instability. Alternatively, policy emphasis on strengthening the stability of the political system tends to reduce resources available for diffusing the growth process more widely. The positive association of the goal of social mobility with greater political stability suggests that variations in achievement with respect to political stability are closely connected with differences in the extent of opportunity to enter the indigenous middle class whose presence typically contributes to a more stable political system. The contribution of the instrument of increased administrative efficiency to greater political stability consists largely of the effect of reduced corruption and nepotism in stabilizing the national government.[19] The other important element on the instrument side of the equation, major investment in expanding transportation and power networks, promotes economic participation over a wider geographic area by providing physical links for enlarging commercial transactions between different geographic regions.

The third and final canonical equation for the high group is also a substitution relationship between economic change and political stability. Specifically, it indicates that, at the margin, for given attainments in enlarging the sphere of modern economic activity, neither the acceleration of overall economic growth as measured by rates of change in real per capita GNP (Y_6) nor major growth in productivity in the agricultural sector (Y_8) can be achieved, on the average, without significant increases in political instability (Y_5):

$$\begin{aligned} & -.48Y_8 - .62Y_6 + .25Y_9 + .48Y_5 \\ & \quad - .03Y_1 + .30Y_3 \\ & = -.46X_8 - .78X_{11} - .01X_3 \\ & \quad + .30X_4 + .23X_{16} + .21X_2 \\ & (\bar{R}^2 = .79) \end{aligned} \tag{17}$$

With respect to the instruments for achieving these goals, the equation suggests that purely economic policies are most effective in both raising growth rates and improving agricultural productivity in countries at the high level of development. In particular, substantial buildup of physical overhead capital (X_8) provides the means for transmitting more widely and accelerating economic growth and technological improvements; while the more rapid improvement of financial institutions (X_{11}) helps to finance the economic expansion of the major sectors of the economy.

In examining the full set of results for countries in the high group, we find that widening participation in modern economic activities and in political institutions is foremost among goals predicatably achievable with available instruments. At the high level, the spread of economic participation is, however, typically achieved at the cost of some significant increase in political instability. The direct raising of economic growth rates as measured by per capita GNP alone is next in importance.

---

[19] These are the elements which weigh heavily in the definition of the indicator.

It should, however, be noted that, in several of the test runs with seven variables, changes in average GNP entered the first canonical function as an important dependent variable associated with purely economic instruments; the significance of this last goal among countries in the high group may therefore well be underestimated by the present set of canonical functions.

VI. *Summary and Conclusion*

Several conclusions may be drawn from the canonical analyses presented in this paper. First, the consistently significant weight attached to reductions in socioeconomic dualism as a goal indicates its primary importance among underdeveloped countries at all three levels of development represented by our data. The fundamental social and economic changes involved in reducing dualistic cleavages in low-income societies are both necessary preconditions to initiate self-sustained economic growth successfully and also have as their characteristic by-product the wider geographical distribution of the material benefits of economic progress.

A second conclusion from our canonical results is the importance of the goal of political stability at all levels of development. This is not surprising since the creation of a stable political framework is a foremost requirement for effective economic development policy in low-income nations in which social unrest and political instability frequently are serious threats to the security and tenure of the leaders. While the goal of increased political stability is significant for all groups, it is particularly crucial for the intermediate-level countries in which, typically, the impact of socioeconomic modernization is extremely uneven and social and political unrest consequently widespread. A further aspect of the findings on political stability is the indication that, at both the intermediate and high levels, policy choices must be made in the allocation of given resources between increases in political stability and those socioeconomic transformations which provide the basis for sustained economic growth.

A third conclusion emerging from our analysis is that the direct raising of real per capita GNP is not foremost among goals dependably associated with available policy instruments at any of the levels of development covered by our data. Our results also indicate the presence of significant substitution relationships between rates of change in average GNP and other critical national goals for all three development levels. For the low level countries, the competing goal is the achievement of socioeconomic and political integration; for the intermediate group, it is the diffusion of the benefits of modern technology throughout the agricultural sector; and for countries at the high level, it is the search for a more stable political framework.

A final conclusion is that in the seventy-four underdeveloped countries represented by our data the spread of popular participation in modern political processes does not usually go hand in hand with the wider distribution of modern forms of economic participation. It is only among countries at the highest of the three levels of development covered in our study that a significant positive relationship appears between increased popular political participation and greater economic integration of disadvantaged groups and sectors.

Broadly speaking, the results of the present canonical analysis are consistent with those which emerged from our earlier factor analytic investigation reported in *Society, Politics and Economic Development.* Specifically, the overwhelming importance at the lowest level of development of the growth of a modern market sector as a stimulus to essential and pervasive social transformations, the great diversity of specific patterns of change both within and between countries characteristic of the intermediate group of countries, and the crucial importance of the wider geographical diffusion of innovations generating growth for the high-level countries, are all findings of the current analysis reflected in our earlier inquiry. The present canonical study adds to the earlier results, however, by specifying the hierarchy of goals which are predictably associated with the policy instruments available to underdeveloped countries. It also suggests the nature and importance of various substitution relationships between competing national goals. It thereby sheds further light on the complicated interactions involved in the sequential processes of social, economic, and political modernization.

APPENDIX A

The Canonical Model

Mathematically, the canonical model hypothesizes the relationship

$$(1)\qquad Y\alpha = X\beta + U$$

Where Y is a $(T\times P)$ matrix of dependent variables; α is a $(P\times 1)$ vector of coefficients to be estimates; X is a $(T\times Q)$ matrix of independent variables; β is $(Q\times 1)$ vector of coefficients to be estimated and U is $(T\times 1)$ vector of normally distributed random disturbances with zero mean, constant variance and no intercorrelation. Note that when Y is a $(T\times 1)$ matrix (and, of course, α is a (1×1) vector), (1) reduces to the simple multiple regression model. As is well known, in the multiple regression model, the least squares estimate of β, $\hat{\beta}$, is

$$(2)\qquad \hat{\beta} = (X'X)^{-1}X'Y$$

and the estimate of R^2, $\hat{R}^2$, is

$$(3)\qquad \begin{aligned}\hat{R}^2 &= \hat{\beta}'X'Y(Y'Y)^{-1}\\ &= Y'X(X'X)^{-1}X'Y(Y'Y)^{-1}.\end{aligned}$$

If we now define

$$(4)\qquad \overset{(T\times 1)}{\tilde{Y}} = \overset{(T\times P)}{Y}\ \overset{(P\times 1)}{\alpha},$$

then by analogy with (2) the least squares estimate of the canonical coefficients β in (1), $\hat{\tilde{\beta}}$, is

$$(5)\qquad \hat{\tilde{\beta}} = (X'X)^{-1}X'\tilde{Y} = (X'X)^{-1}X'Y,\hat{\alpha}$$

where $\hat{\alpha}$ is the least squares estimate of α. To obtain $\hat{\alpha}$, we maximize $\hat{\tilde{R}}^2$, the square of the canonical correlation. This is of course equivalent to the minimization of residuals (the usual least squares criterion). By analogy with (3) we write

$$(6)\qquad \begin{aligned}\hat{\tilde{R}} &= \tilde{Y}'X(X'X)^{-1}X'\tilde{Y}(\tilde{Y}'\tilde{Y})^{-1}\\ &= \alpha'Y'X(X'Y)^{-1}X'Y\alpha(\alpha'Y'Y\alpha)^{-1}.\end{aligned}$$

To maximize $\hat{\tilde{R}}$ we differentiate (6) with respect to α; maximization with respect to β has already taken place through the substitution of $\hat{\beta}$ in the definition of $\hat{\tilde{R}}^2$ in (3). The maximization of $\hat{\tilde{R}}$ is carried out under the constraints that the variances of $\tilde{Y}$ and of $X\beta$ each be unity in order to obtain a unique solution. We therefore require that

$$(7)\qquad E(\tilde{Y}'\tilde{Y}) = E(\alpha'Y'Y\alpha) = 1$$

and

$$(8)\qquad E(\beta'X'X\beta) = 1.$$

To maximize $\hat{\tilde{R}}$ with respect to α subject to constraints (7) and (8) we now form the Lagrangian equation

$$(9)\qquad \begin{aligned}V = {}&\alpha'Y'X(X'X)^{-1}X'Y\alpha\\ &- 1/2\lambda(\alpha'Y'Y\alpha - 1)\\ &- 1/2\mu(\beta'X'X\beta - 1)\end{aligned}$$

and differentiate it partically with respect to α, λ and μ. In (9) the first term is not postmultipled by $(\alpha'Y'Y\alpha)^{-1}$ as in (6) since the expected value of this term is unity by (7).

The differentiation yields

$$\frac{\delta V}{\delta \alpha} = Y'X(X'X)^{-1}X'Y\alpha - \lambda Y'Y\alpha = 0 \tag{10}$$

$$\frac{\delta V}{\delta \lambda} = \alpha' Y' Y\alpha - 1 = 0 \tag{11}$$

and

$$\frac{\delta V}{\delta \mu} = \beta' X' X\beta - 1 = 0 \tag{12}$$

Equation (10) can be rewritten as

$$[Y'X(X'X)^{-1}X'Y - \lambda(Y'Y)]\alpha = 0 \tag{11}$$

which can only be satisfied if

$$| Y'X(X'X)^{-1}Y'Y - \lambda(Y'Y) | = 0 \tag{12}$$

The solutions to characteristic equation (11) are the eigenvalues λ and the associated eigenvectors α. There are, of course, as many eigenvalues and eigenvectors as the dimensions of α (i.e., P). The corresponding estimates of $\hat{\beta}$ is obtained by substituting $\hat{\alpha}$ in (5).

The roots λ are the squared values of the canonical correlations as can be seen by rewriting (12) as

$$Y'X(X'X)^{-1}X'Y = \lambda(Y'Y) \tag{13}$$

Postmultiplying (13) by $(Y'Y)^{-1}$ yields

$$Y'X(X'X)^{-1}X'Y(Y'Y)^{-1} = \lambda \tag{14}$$

which is identical to (3), the definitional equation for $\hat{R}^2$.

The successive roots are in descending order of magnitude, since they are extracted by maximizing $\hat{R}^2$; the successive eigenvectors are, of course, orthogonal to each other. One may, therefore, think of the successive orthogonal triplets $(\lambda, \hat{\alpha}, \hat{\beta})$ as defining hierarchies of independent dependence relationships.

APPENDIX B
Structural Models for Selected Target Variables

Model for Dualism and Social Mobility. This model consists of three equations. According to the first one, the widening of the base of modern economic activity (represented by positive changes in dualism, Y_9) is directly associated with the creation of a nationwide transportation and power network (X_8), the emergence of a class of indigenous entrepreneurs and traders (Z_1), the acquisition of literacy (X_1), and the enlargement of economic opportunities in the foreign trade related modern sector of the economy, represented by the indicator of the structure of foreign trade (X_{15}).

$$Y_9 = .47X_8 + .13Z_1 + .13X_{15} + .20X_1 \qquad (R^2 = .89) \tag{18}$$

In this equation all variables except for Z_1 appear in our list of instrumental variables.[20] To round out the model, we therefore need an endogenous explanation for intercountry differences in the size of the indigenous middle class.

The equation which is specified for the indigenous middle class in the present model suggests that the growth of an indigenous middle class depends upon the more generalized access to economic opportunities provided by increased social mobility (Y_1) and upon the provision of finance for the expansion of indigenous entrepreneurial activity, summarized by our index of the rate of improvement in financial institutions (X_{11}).

$$Z_1 = 21.68 + .75Y_1 + .22X_{11} \qquad (R^2 = .78) \tag{19}$$

---

[20] The indicators of the importance of the indigenous middle class (Z_1) and of the political strength of the traditional elite (Z_2) are variables appearing in the present structural model which are not included in the canonical analysis because they are not clear-cut cases of either goals or instruments.

The extent of social mobility is endogenous in this relationship and must therefore be "explained." The equation derived to explain it in our model includes several diverse, interrelated forces: the spread of the capacity for social communication through literacy (X_1), the provision of more "room at the top" associated with the enlargement of the indigenous middle class (Z_1), and the weakening of political barriers to social advancement represented by a reduction in the political strength of the traditional elite (Z_2).

$$Y_1 = .53X_1 + .37Z_1 - .20Z_2 \qquad (R^2 = .88) \tag{20}$$

If Z_2 is regarded as predetermined, the three equations presented above can be solved simultaneously to obtain the reduced form equations for dualism and social mobility, respectively. As already given in the text, these are:

$$Y_9 = 3.93 + .47X_8 + .27X_1 + .131X_{15} \tag{3}$$

and

$$Y_1 = 11.10 + .734X_1 - .076Z_2 + .113X_{11} \tag{4}$$

Model for Level of Agricultural Productivity. The forces that are significant in explaining differences in the level of productivity in agriculture are those which impinge upon the productivity of its human resources (literacy, X_1), possibilities for participation in the market associated with the generalization of economic growth (dualism, X_9), and major increases in the provision of feeder roads to agriculture (X_8).

$$Y_8 = .52X_8 + .28Y_9 + .20\, X_1 \qquad (R^2 = .83) \tag{21}$$

This equation can be solved together with the reduced-form equation for dualism (3) to yield the reduced-form equation for the level of agricultural productivity which was presented in the text:

$$Y_8 = 1.12 + .66X_8 + .28X_1 + .04X_{15} + .01X_{11} - .01Z_2 \tag{5}$$

Model for Size of the Traditional Agricultural Sector. The important influences tending to reduce the extent of subsistence agriculture in the present structural model are the geographic widening of the area of economic growth stimulated by the rise of a modern sector (dualism, Y_9), the physical linking by transportation facilities of the modern sector with the hinterland (X_8), and the increased psychic contact between agricultural and urban populations made possible by the growth of literacy (X_1).

$$Y_7 = -.28Y_9 - .35X_8 - .25X_1 + 95.6 \qquad (R^2 = .85) \tag{22}$$

This equation can be solved together with the reduced form equation for dualism (3) to yield the reduced-form equation already given in the text:

$$Y_7 = 96.71 - .33X_1 - .48X_8 - .04X_{15} - .01X_{11} - .01Z_2 \tag{6}$$

CENTRAL LIBRARY. UNIVERSITY

DISCUSSION

ROBERT E. BALDWIN: As suggested by the Program Committee, I shall devote my time as a discussant to the paper by Professors Fei and Ranis.

One of the admirable characteristics of the excellent work that Professors Fei and Ranis have done in the development field over the past several years is that their studies always remain solidly focused upon the central question of development economics; namely, what are the major forces and mechanisms that determine an economy's rate of growth? In the process of constructing elaborate growth models, it is all too easy to become so intrigued by the various logical possibilities of the model that one ends up with an analytical exercise that has little to do with the main question we are interested in. Fortunately, Fei and Ranis never engage in trivial theorizing of this sort, but instead keep asking themselves how their analysis bears on the central issues of the subject. The paper they have given today is an example of this approach.

There are, it seems to me, two major points that stand out in their sweeping survey of the development process. The first is that the existence of what they call "intersectoral connectivity" serves as a major growth-propelling force. The second is that development history over the last four hundred years can be usefully viewed as a process that involved changes in the quality and quantity of capital accumulated by an economy. Both of these ideas can—as the authors point out—be traced back to early writers in economics, especially to Adam Smith.

In reviewing their application of these concepts to the history of economic development since the sixteenth century, one might stress the standard objections to adopting a stages-approach to economic history or cite the dangers of stating that one economic system inevitably leads to another. However, I shall not take this tack here. It is most useful to emphasize such factors only when some simple stages-model is on the verge of widespread acceptance. However, the points made by Fei and Ranis, in my view, not only deserve much wider acceptance than they have, but they do not lead to any simple stages-theory of the development process. Consequently, I shall mainly elaborate upon the major concepts in their paper.

The notion that the interactions and feedbacks resulting from the creation of exchange relationships between economic sectors are a major source of growth is one that development economists have rediscovered only comparatively recently. Most neoclassical economists, for example, assumed these relationships to be in existence already, except in the international sector, and they, consequently, concentrated on other growth factors. Yet, it is increasingly apparent from a study of the current less developed countries—and the authors point out the consistency of this view with the economic history of the developed countries—that the creation of extensive trading relationships among sectors, especially between rural and urban groups, is not only essential for socially satisfactory economic development but is a powerful growth-creating force.

The latter point follows, first, from the fact that most rural and urban producers seem to be highly responsive to economic incentives, especially when given an opportunity to increase the variety of commodities they can consume. Second, most rural and urban producers already possess the knowledge and capital to raise their productivity levels somewhat in response to improved market opportunities by means of greater specialization and a higher degree of capacity utilization. Finally, they are able to raise productivity even more because of the spread of technical knowledge stemming from the greater contact among individuals that is associated with increased trade. The resultant growth in income permits still more capital accumulation and further specialization.

This description of the development process is, of course, nothing more than Adam Smith's description of cumulative growth. What Smith calls man's natural propensity "to truck, barter and exchange one thing for the other" is what modern development economists have found to be the high degree of responsiveness on the part of producers and workers in the less developed countries to economic incentives. The rest of the description is simply his statement of the interaction between greater division of labor and specialization and capital accumulation.

Although one must go back to Smith to find the greatest emphasis by an early writer on the importance of sectoral interactions as a growth mechanism, the notion did not entirely disappear from economics after Adam Smith. It still continued in the concept of internal and external economies as expounded by such writers as Marshall and Allyn Young. In particular, these writers recognized that in a developing economy pecuniary spillovers often produce more than merely a reallocation of existing resources. They also lead to an increase in the productive powers of these resources for the kind of reasons cited by Adam Smith.

Unfortunately, as neoclassical economics turned more and more to static analysis, the notion of production externalities was turned into a trivial concept and pushed into the background of economics. It has only been in relatively recent years that such authors as Scitovsky and Rosenstein-Rodan have again put the concept in its original dynamic setting. The notion of backward and forward linkages stressed by Hirschman as well as the intersectoral analysis of Fei and Ranis are other modern manifestations of this basic concept.

It is useful to think of sectoral interrelationships in broad terms and divide them into three types: those among different geographical areas; those among different industries; and those among different income groups. As Fei and Ranis point out, the central feature of mercantile agrarianism is the interregional and international spread of economic relations through trade in agricultural products. Industrial capitalism, on the other hand, involves not only interregional trade but trade in widely different types of products; namely, agricultural and manufactured commodities. Furthermore, both economic systems involve economic interrelationships among distinct income groups. In mercantile agrarianism the groups are the capitalists, workers, and landlords whereas in industrial capitalism capitalists and workers become the two increasingly important income groups as urban areas absorb a growing proportion of the population.

It seems to me to be useful to stress the growth-generating aspects of the economic interrelationships among income groups somewhat more strongly than the authors do in their present paper. For example, the driving force of the Ricardian or Marxian models is not geographic or interindustry interrelationships but rather the economic interactions among income groups. In particular, the effect of capital accumulation in setting off repercussions that eventually reduce the upward pressure on wages and thereby tend to keep profit rates up is essential to both models. In the Ricardian case the downward wage pressure comes from the population growth stimulated by capital accumulation; whereas in the Marxian case it is the laborsaving nature of the accumulation process that prevents profit rates and growth from rapidly converging to near zero levels.

The pattern of income distribution in an economy is important for development at all stages of growth but it becomes especially significant once an economy becomes completely commercialized. There are a number of ways in which it can affect the rate of growth. One set of repercussions stems directly from the nature of consumption demands and savings that a given income distribution generates. For some less developed economies of the past and present, the problem seems to be not so much the lack of an adequate star-tree structure—to use the Fei-Ranis term—as the existence of an income distribution that has a low growth-stimulating effect. Still another way in which development is affected by the distribution of income is through the social and political reactions that generally seem to be associated with various income patterns. One of Marx's main contributions was to stress how a society could be torn apart by a highly unequal allocation of income. In our own society—what Fei and Ranis call technocratic capitalism—we see that there is an important part of the labor force, namely, unskilled labor, that is increasingly being rendered obsolete. Unless this group is permitted to share in the benefits of development in a productive manner, the maintenance of a satisfactory development rate will be jeopardized on social and political grounds. Still another example of the significance of income distribution is the political and social pressures against economic change that frequently go hand in hand with the plantation-type economy.

Given a development framework that views the major source of growth as coming from the interactions among regions, industries, and income groups, such traditional motivating factors as population growth and technological progress become in large part endogenous rather than exogenous variables. A good example is the rate of technical progress in modern advanced economies. By means of such institutional forms as the giant corporation and various types of state subsidization of other organizations, the modern technocratic economy has been able to make the rate of technical progress a variable that responds to and reacts with other endogenous variables. Writers like Schumpeter and Marshall already saw that rapid technical progress was largely a given factor under industrial capitalism. Now it can be viewed as one of many variables determined by the nature of intersectoral repercussions.

The quantity and quality of capital accumulated by an economy has long been regarded as determined in part by the economic interactions among groups and sectors and in part by the rate of technological progress and by natural resource conditions. In the classical model, as Fei and Ranis point out, the capital stock ultimately depends upon the size of the population—a factor that itself is determined within the model. In the Marxian and neoclassical models it is tied to the rate of technical progress and the nature of saving propensities, since it consists largely of equipment and structures rather than a stock of consumption commodities. In the technocratic economy of today, the rate of capital accumulation is once again becoming related to the size of the population. Now the key form of capital accumulation is investment in human resources. In this connection it is interesting that as the rate of technical progress as well as the pace of fixed-capital accumulation became increasingly internalized within the large corporation, the nature of technology began to create a demand for a type of capital accumulation outside of the control of the corporation. This, of course, was a demand for highly educated individuals both to cooperate in the productive process and to generate new technology. Because of the special externality problems associated with educational investment as well as other forms of investment in human resources, education—even at the highest levels—has been undertaken largely outside of the private profit framework. A significant part of research activities also seems to be falling within this category. Thus, a major problem that technocratic capitalism now faces is how to control the interactions among the public and private sectors so as to produce not only a rapid rate of growth but a growth pattern that brings significant income benefits to all groups. As Marx and Schumpeter brought out, what matters is not just the rate of growth but the distribution pattern associated with it. Valuable insights into the nature of the solution to this problem as well as to the development problems faced by less developed countries can, I think, be achieved by exploring in more analytical and empirical detail the concept of intersectoral interactions that is the key feature of the Fei-Ranis growth model.

CHARLES R. FRANK, JR.: First, let us consider the paper by Adelman, Geier, and Morris. Mrs. Adelman and her colleagues have performed an important service to the economic profession in compiling a considerable amount of data and indices for a large number (seventy-four) of countries and in running their cross-country factor analyses and canonical correlations. The paper, however, consistently neglects the importance of having a well-articulated theory before applying empirical tests. While it is true that evidence of association and correlation among sets of data may be suggestive of possible theoretical relationships, inductive theorizing of this sort is very dangerous and involves many more caveats than the authors seem willing to admit. In many instances there are a number of alternative, more simple, and more intuitively plausible explanations of their results which often could have been used to predict the results beforehand.

To be more specific, let me describe in detail a number of objections. First, the paper is an outstanding example of what a former mentor of mine calls "correlation hunting." Mrs. Adelman and her colleagues have available a large number of possible instrument variables and goal variables. From these a smaller and equal number of instruments and goals are chosen. The criterion used is the size of the simple correlation coefficients between instruments and goals. Instruments may then be added to this basic set if any one individually adds significantly to the canonical correlation.

In deriving the so-called "structural" equations, for each dependent variable the authors specify a "set of potentially causal independent variables." The variables actually included in the "structural relationship" then are those which pass a stepwise regression test of significance.

There simply are no justifications for any of these procedures. If their purpose is to produce high canonical correlations or high multiple correlation coefficients in the structural equations, then the high correlations they obtain have extremely little statistical significance. Given the large number of variables with which they are working, pure chance alone should produce some fairly high correlations and some seemingly significant variables. In addition, it is most disconcerting to see the authors emphasize particular variables as highly significant when the decision rules used to include variables in a relationship may exclude other variables which are also highly significant. More precisely, there are three reasons why this is so: (1) a variable with a near-zero simple correlation coefficient with every other variable may be highly significant when entered into a multivariate regression analysis; (2) a variable which fails to pass a significance test in a later stage of a stepwise regression procedure may indeed pass the significance test if entered as one of the original variables in the relationship or at an earlier stage in the stepwise procedure; and (3) variables A and B tested singly in a stepwise regression procedure may not prove significant but introduced together, either or both may prove significant.

In the first lecture to students of development economics, one often warns them to be wary of how statistics on less developed countries are compiled. For example, data on food output are often determined by applying figures on average farm food production derived from out-of-date sample surveys to rural population data. Thus, if one tries to regress food production on rural population, one should not be surprised at the high correlation; the regression coefficient is not a parameter of a production function but an estimate of the average farm production figure used by the statistician to arrive at his total food production data. Similarly, Mrs. Adelman and her colleagues produce for us good fitting relations in which roughly the same basis is used for constructing the values of both dependent and independent variables. Since they themselves manufactured their data, surely they should not expect us to take these regressions seriously. For example, we are told that the importance of "an indigenous middle class (Z_1) depends upon the more generalized access to economic opportunities provided by social mobility (Y_1)." The high correlation, of course, is not surprising given that the value of the variable, social mobility (Y_1), consists of two main elements, one of which is "measured . . . by the importance of the indigenous middle class." (See Adelman and Morris, *Society, Politics, and Economic Development*, 1967, p. 16.) Lower down, the direction of the interdependence between Z_1 and Y_1 is reversed and not too surprisingly another significant relationship emerges. This kind of analysis appears again and again. Thus, the level of agricultural productivity (Y_8) and the relative size of the traditional agricultural sector (Y_7) are regressed on dualism (Y_9); the first (Y_8) is measured partly by "the relative weights of traditional and modern agriculture" (*Society, Politics, and Economic Development*, p. 104) and the last (Y_9) partly by the existence and size of the modern production sector relative to the traditional sector. Obviously, we find a high degree of correlation. In the present paper we find, as one might expect, that the extent of dualism (Y_9) and the character of agriculture organization (X_{16}), which are both measured in part by the pervasiveness of owner-absentee commercial farming, are closely related.

Another difficulty encountered in the paper is the interpretation of causality placed on the canonical regressions. Many of the instrument variables and goal variables are well-known concommitants of economic development, and it is not surprising that they are closely related to each other. The fact that these variables correlate in a statistical sense in no way establishes the strength or direction of causality. In fact, I would argue that many of the so-called "goals" are necessary consequences of the process of economic development and these consequences themselves induce changes in the values of the so-called "instruments." For example, the extent of social mobility (Y_1), which is measured as a combination of the ratio of relevant population in primary and secondary school and the proportion of the working population in "middle class" (i.e., skilled) occupations, is inevitably associated positively with the level of economic development. Thus, it is not surprising to find in the first canonical relationship that social mobility and two other goals, the size of the traditional agricultural sector (Y_7) and the extent of socioeconomic dualism (Y_9), are all positively related to the two instruments, adequacy of physical overhead capital (X_8) and the rate of literacy (X_1). All of these variables are most surely associated positively with the level of economic development. The high correlation is not necessarily indicative of causality; in fact, the most obvious causality runs from the goal, social mobility (which includes a measure of the proportion of school-age children receiving an education), to the instrument, the rate of literacy, rather than the other way around.

Similarly, the positive relationship in the second canonical correlation between rate of growth of GNP (Y_6) and the index of improvements in financial institutions (X_{11}) is not surprising and the direction of causality is not obvious. The authors assume, as they do throughout the paper, that changes in the value of the goal variables are caused by movements in an instrumental variable; i.e., "financial improvements accelerate economic growth." The variable financial improvements (X_{11}), however, is a combination of the percentage point change in the ratio of time and demand deposits to GNP and the increase in the real value of private domestic liabilities to the banking system. One would expect economic growth to be accompanied by increasing monetization of transactions and a decreasing reliance on cash for transactions. The amount of increase in the real value of private domestic liabilities would normally be associated with an increase in the real money supply which in turn is certainly related positively to growth in real GNP. Viewed in this way, an increase in the so-called "index" of improvement in financial institutions is an expected consequence of rapid growth in GNP rather than a "cause" of economic growth.

Despite my lack of sympathy with their methodology of model building and the conclusions drawn from their empirical analysis, the paper by Adelman, Geier, and Morris must be regarded as a brave attempt to synthesize political, social, and economic approaches to development through empirical analysis. Their use of canonical correlation analysis is a first attempt to analyze a considerable amount of cross-country data in this fashion. Their paper, however, does not seem to be very successful, mainly because the theoretical underpinnings do not exist. I suspect that before empirical analysis of this sort can be a very useful exercise we will require much more effort in developing models of social, political, and economic change which include meaningful analysis of the interactions among the economic and noneconomic variables.

The first two-thirds of the Fei-Ranis paper are a fascinating account of the transitions from agrarian to mercantile to industrial economies. Their explanation of the role of capital in the three stages is particularly instructive. The presentation is somewhat marred, however, by frequent references to the so-called "tree-star" structure which are rather unnecessary and unhelpful intrusions into the main line of thought.

Although I found their historical account quite fascinating, the attempt to apply the lessons of history to modern less developed countries has a decidedly forced tone. Their assumption that, relative to mercantilist societies, the less developed countries of today are handicapped by lack of social overhead capital is not very satisfactory. Road networks in most less developed countries today are highly developed relative to seventeenth and eighteenth century systems in Western Europe. More importantly, the LDC's today have bicycles, motor vehicles, railways, telephone, telegraph, radio, and television communications which were unheard of two and three centuries ago. The facilities for trade and "contact" in any important sense in less developed countries today are far superior to those of the Western world during the mercantile period. The reasons for the stagnation of trade in some less developed countries must be sought elsewhere. We have had too many experiences in the last two decades with grandiose schemes for improving social overheads which have come to naught. It is about time we recognized this and shed our more simplistic notions about the usefulness of social overhead capital.

The paper by Lloyd Reynolds on the content of development economics is an extremely useful catalogue of topics of interest to present and prospective development economists. My only slight quarrel with the presentation derives from a lack of faith on my part in the ability to derive a "theory of economic development" along either neoclassical or neo-Keynesian lines. It seems to me that a most important contribution would be a theory which concentrates on the importance of "agents of change" or entrepreneurs along Schumpeterian or Hirschman lines. We need desperately a really successful integration of social, political, and economic variables into some satisfactory theory of social and economic change.

CARLOS F. DÍAZ-ALEJANDRO: Lloyd Reynold's general propositions on the nature and content of development economics strike me as sensible; so my remarks on his paper will deal with some aspects of research strategy aimed at better understanding the operation of LDC's economies.

Aggregate quantitative analysis using both time series and cross-section data is becoming easier with each post-1945 year passing without a major war and with each new publication of a national or international agency. A standard excuse for shunning quantitative research in the development field ("data are not there") is losing credibility. But the lure of readily available United Nations yearbooks may be inducing a relative excess of studies using aggregate variables.

Few would deny that aggregation problems are more acute in LDC's than in DC's, making macroeconomic variables more suspect in the former (independently of the quality of basic statistics or of statistical offices). The conceptual problems multiply when making international comparisons of such variables; among other things, relative price structures can differ very sharply (more so than between Europe and the U.S.) from one LDC to another, depending on how much public policy has isolated one economy from world markets. Comparisons between such variables as savings rates, shares of manufacturing in GNP, etc., lose much meaning unless adjusted for relative price structures. Although much useful work has been done using aggregate variables right out of UN yearbooks, their future expansion faces sharply diminishing returns.

Even when aggregate variables can be defined unambiguously, some of the comparative and cross-sectional studies suggested by Reynolds are likely to yield neither a better understanding of the operation of LDC's economies nor lay a firmer basis for policy making in those countries. Suppose research shows that, correcting for differences in relative prices, government output as a percentage of GNP is positively related to a country's per capita income or to its population. This will be interesting, but the inevitable "so what" will be hard to answer. At best, a few antigovernment extremists may be rattled. Or suppose one showed, as indeed one could, that during the postwar years Latin-American countries with the fastest growth in per capita GNP have also tended to have the highest rates of population expansion. This may confound the most ardent advocates of population control, but it will add no insight into the connections between the two variables.

Because of the ambiguities of aggregate variables for LDC's and the sterility of many aggregate cross-country generalizations, I hope Reynolds is right in predicting that the largest kind of research activity will be microanalysis. If we are not to throw away our tool kit when examining LDC's economies, then a major task is to examine at a low level of aggregation in which way standard economic parameters, such as price elasticities of various kinds, differ between DC's and LDC's, and among LDC's. Even in areas where aggregate analysis is more appropriate (e.g., the effect of foreign aid and direct investment on the receiving country) a retreat to a more modest scope is likely to provide insights now lost in gaseous generalizations.

Quite properly, however, development economists will continue doing whatever each thinks is most fun. And the irresistable fun of reaching for high theory and magnificent dynamics is well illustrated by the stimulating Fei-Ranis paper. So after exhorting the profession to think small, I would like to briefly join Fei-Ranis in their romp through history—a bit of "economists' imperialism" likely to leave learned historians reeling. There is much in their paper with which I agree, but I will center my remarks on disagreements.

Fei-Ranis overemphasize capital formation and underplay technological change in discussing forces promoting secular growth. Taking note of recent studies on the sources of growth in advanced economies, they admit the importance of the search for applied scientific knowledge, but only in the twentieth century. We do not have much quantitative information for earlier centuries, but I would hypothesize that throughout history any observed deepening of capital has been possible mainly through the advance of technical progress, which held up the rate of return. Compare the Roman Empire of the second century with Western Europe of the eighteenth century. What did the latter have which the former did not? The major difference, I suspect, was not in surplus above minimum consumption, nor in the wider trading network of Western Europe compared with the Mare Nostrum common market. The major difference rested in the fund of technical and scientific knowledge, and in its steady increase, the lack of which doomed Latin Power (in spite of its well-developed cities, roads and markets—its "tree-star structure").

The limited way in which Fei-Ranis take into consideration technological change is illustrated by the transition from simple to mercantile agrarianism. Trade opens up, as suddenly as in the pedagogic exercises of international trade theory, without reference to scientific and technological changes in navigation and shipbuilding, nor to other technological changes explaining the original creation of the trade margin. In their model, agricultural technological change is made to depend on the expansion of mercantile capital. There is insight in this hypothesis, but it neglects other kinds of technical change which were causes rather than effects of trade.

While a given agricultural revolution may occur simply by narrowing the gap between average and best technology within a country, the social rate of return on investment in leading countries would quickly (by our historical standards) drop very low without a continuing improvement of best technology. The stagnation and decline of great mercantile empires of the past may be better explained by their failure to continuously improve best technology, above and beyond those changes induced by mercantile expansion, than by their lack of capital in the Fei-Ranis sense.

The emphasis on capital formation, fed by the surplus above minimum consumption and inducing technical change, leads Fei-Ranis to an overly optimistic view of historical progress. This view ignores not only Roman decline, but also that of other advanced mercantile societies, such as medieval Florence and Venice. (Incidentally, the Fei-Ranis description of pre-1500 Western Europe involves a simplification of heroic proportions.) In several of those societies, the factors listed by Fei-Ranis as inducing a favorable interaction between trade and productivity were present, but failed to lead to industrial capitalism.

Their view that, both under simple and mercantile agrarianism, growth in population pushed is based on the assumption of an exogenous positive population growth rate, an assumption which does not only go against Malthus, but also against history (e.g., the Black Death). They also slight the classical economists' historically motivated preoccupation with land. Especially under agrarianism, land ownership plus the technology of rural activities (e.g., crops versus livestock) had a greater social impact than control of working capital.

Passing from historical review to their conclusions regarding its relevance to present-day LDC's, I also have some quibbles. Is the concept of tree-star structure operational? Can it by itself differentiate between a road project with a high economic payoff and wild schemes to open up the hinterland? In regions like Latin America where distances are enormous in comparison with those of Western Europe, the costs of developing a "thick" tree-star structure can become very large. And I fear one can easily imagine many a politician trying to justify his pet infrastructure scheme by references to trees and stars.

Few would quarrel nowadays with the need to differentiate among LDC's and to stress their heterogeneity. But how far a growth-theoretic typology, advocated both by Reynolds and Fei-Ranis, should be developed is another matter. After all, in spite of structural differences, few argue the need to develop different economic theories for Holland and the U.S.A. It may be better in the long run to approach the variety of LDC's via the quantification of standard parameters (such as the elasticity of land supply, import propensities, etc.) rather than by developing models which by their nature would tend to exaggerate differences and place undue emphasis on such things as the colonial past.

After criticizing the Fei-Ranis paper for not solving the puzzle of Roman decline and fall and for not having unraveled all the mysteries of inventions and technolgoical change, I will add that their paper fully succeeds in its aim "to probe and stimulate." I also share their faith in the fruitfulness of economic history for enriching our understanding of development.

PROBLEMS IN THE AREA OF POVERTY

THE OBE SIZE DISTRIBUTION SERIES: METHODS AND TENTATIVE RESULTS FOR 1964

By EDWARD C. BUDD, *Pennsylvania State University*
and
DANIEL B. RADNER, *Office of Business Economics*

If this paper can properly be said to be concerned with problems of poverty, it is in the area of measurement and of data needed to implement both the measurement of poverty and the analysis of its causes. Measuring the number of poor requires not only a definition of the poor but an income size distribution series with sufficient demographic and economic characteristics to implement the definition. The current one requires information on such characteristics as family size and composition and farm-nonfarm residence. But a count of the poor is not the economist's only concern: he needs data to assist him in diagnosing the sources of low incomes and in measuring their quantitative importance. Hopefully, a new series would help fill both of these needs in poverty research.

The scope of the research we are reporting on, however, encompasses more than the area of poverty, since it is concerned with all parts of the income distribution, not simply those at the bottom. Indeed, insofar as poverty is thought of as something relative—as the status of income recipients in the lower part of the distribution relative to the positions of those above them—the entire distribution becomes relevant to its measurement.

The government data files that have been available for research in poverty and in income distribution generally, either in the form of published tables or in some cases the underlying computer tapes furnished to users, have come from the decennial Census or from field surveys, such as the March income supplement of the Census's Current Population Survey (CPS), the BLS's Survey of Consumer Expenditure, and the 1962 Survey of Financial Characteristics of Consumers (SFCC) conducted by the Census for the Federal Reserve Board. Soon, we hope, the OEO's Survey of Economic Opportunity for 1965 and 1966 will be added to the list. Without the existence and availability of these files, much of the research and measurement in the area of low incomes could not have been carried out.

This does not mean that the data obtainable from existing field surveys are free from deficiencies. For one thing, they may not provide the kind of information needed for the study of certain problems. For example, it is impossible with existing sources to study the behavior of income and income recipients over more than one or two years, or to take account of other than a conventional list of demographic and economic variables. Entirely new surveys may be required to fill the gap. One such project—OEO's longitudinal study—is the subject of the next paper at this session.

Other well-known problems arise from response error and nonresponse. These are particularly serious for the income data: the results are the underreporting of total income and its components and the rather uncertain relationship that may exist between the reported income of the recipient unit and its "true" or actual income. No measure of the latter exists, of course, and the degree of correspondence must be judged by comparing what the same unit reports in two or more different sources;

e.g., the CPS and Census or Census and IRS. The lack of consistency in reporting which matching studies show, particularly for certain income types, is disquieting indeed.

Estimates of the extent of underreporting of total income in field surveys vary from about 15 percent in the CPS to 6 to 9 percent for the decennial Census, and 9 percent for the SFCC, when OBE's family personal income, adjusted for conceptual comparability with the money income concept employed in these surveys, is used as the control. For some individual income types differences between the amount reported and the control are even greater, amounting in the March, 1965, CPS (income year 1964) to as little as 2 percent for nonfarm self-employment income and 8 percent for wages and salaries to 41 percent for farm income and 46 percent for other income. Differences in the components of other income are even more striking: social security and railroad retirement were 87 percent of the control; public assistance and unemployment compensation, 64 percent; property income (rent, interest, and dividends), only 37 percent; and all other (inclusive of such items as pensions, annuities, and veterans' benefits), 55 percent.

These differences arise, not simply because the respondents on the average understate their incomes, but because some of them fail to report on certain types they in fact have, particularly property income. One test is provided by a comparison of CPS with tax returns for 1964. While 43 percent of all returns filed reported the receipt of some property income, only 23 percent of roughly comparable CPS units did so. For joint returns and married couples, most of whom can be presumed to have filed returns, the percentages are 49 and 33, respectively. Of course, the tax return reporting rate for interest and dividends itself increased sharply in the years immediately following the initiation of reporting by financial and other institutions to IRS of interest and dividends paid to individuals. For interest, the reporting rate rose from 16 percent in 1961 to 34 percent in 1964; for dividends, from 11 percent to 15 percent. Even these latter percentages may be too low, judging by data from the SFCC, where a more concerted effort was made to obtain property income.

Insufficient research has been done on problems of errors in response by those who do answer the income questions to permit us to judge fully the reasons for the discrepancies noted. A number of possibilities have been suggested for the CPS: the person with the most knowledge of the household's income, presumably the head, is not typically the one interviewed; the household is not advised before the interview of the questions to be asked, and is not asked to consult its records; the survey is conducted a month before persons are required to file their tax returns; the questions on previous year's work experience, which would serve as a reminder on earnings, are not asked in the same interview as the income questions; the income questions themselves may not be sufficiently well phrased or sufficiently detailed or the interviewers sufficiently skilled. Census is currently considering proposals for improving the March income supplement and for engaging in research to solve some of the problems mentioned above. But it would be overoptimistic, we believe, to count on these measures to solve the problems completely, and other methods of improving the income data need to be explored.

Another possible source of downward bias in the CPS comes from the failure of many of those interviewed—9 percent of all persons in the March, 1965, CPS, even higher in subsequent years—to answer the income questions. Under current Census procedures, nonrespondents (so-called "NA's") are assigned income amounts and types based on an actual respondent with similar (selected) demographic and economic characteristics, or, if no such respondent can be found, certain predetermined average values based on such characteristics. How to adjust the sample for NA's is one of the most difficult problems faced in the CPS. While Census techniques for imputing incomes to NA's probably have not been able to take full account of the presumed relationship between nonresponse and income, it is difficult to devise adjustment methods that will. At OBE, for example, we experimented with a considerably more detailed matrix of characteristics for allocating respondent's incomes to NA's in the March, 1965, survey, employing some 1,500 cells, compared with the less than 300 cells in the Census matrix. While the relationship between nonrespondents' economic and demographic characteristics and the incomes assigned to them was undoubtedly improved, the income total allocated was, if anything, less than that originally allocated by Census. For married couples, for example, the average income allocated to nonrespondents by OBE was $1,160 more than the average income of respondents, although $120 less than the corresponding Census allocation. Similarly, improved allocation procedures developed by the Census in connection with the introduction of the 1107 computer resulted in the assignment of a somewhat lower median income to male nonrespondents, and about the same, to female nonrespondents, as compared with their old methods, for the income year 1966.

It may be useful to get some rough idea of the effects of underreporting of income types on size distribution and on the measurement of poverty. We have tried to provide such a measure by first tabulating the data based on incomes as originally reported in the CPS; then blowing up each income type for each record by the ratio of the OBE control total for that type to the total amount actually reported in the CPS, and retabulating the results. Table 1 gives the results for the size distribution of total money income for fami-

TABLE 1

Distribution of Money Income in 1964 for Consumer Units Based on the March, 1965, Current Population Survey
(Millions of units)

Total Money Income	Families*		Unrelated Individuals	
	Original	Adjusted	Original	Adjusted
Under $ 1,250	2.83	2.01	4.32	3.03
$ 1,250– 2,749	5.83	4.33	2.94	3.34
2,750– 4,749	8.34	7.55	2.19	2.24
4,750– 6,749	9.69	8.92	1.46	1.62
6,750– 9,749	11.42	11.48	.75	1.09
9,750– 14,749	7.98	10.09	.25	.47
14,750– 24,749	2.47	3.79	.09	.18
24,750 and over	.55	.96	.04	.09

* The sum of families and subfamilies as defined by Census.

"Original": Based on tabulations from the March, 1965, CPS, reweighted and nonrespondents allocated by OBE.

"Adjusted": Same, after each income type was inflated to the OBE control total for money income.

SOURCE: Office of Business Economics.

TABLE 2

CONSUMER UNITS ABOVE AND BELOW POVERTY IN 1964, BASED ON THE CURRENT POPULATION SURVEY
(Millions of units)

	Families*		Unrelated Individuals	
	Original	Adjusted	Original	Adjusted
In poverty..............	7.60	5.52	5.04	3.83
Not in poverty..........	41.52	43.60	7.01	8.22

* The sum of families and subfamilies as defined by Census.

Definitions and source: Same as Table 1. Poverty line based on the Social Security Administration's poverty index, from Mollie Orshansky, "Recounting the Poor—A Five Year Review," *Soc. Sec. Bul.*, Apr., 1966.

lies and unrelated individuals. Frequencies in the lower brackets are reduced, and those in the upper brackets increased; for families, the number below $2,750 is reduced by a third, and the number with $14,750 or more is increased by three-fifths. In Table 2 for poverty levels, the effect is to reduce the number below the Social Security Administration's poverty line by 27 percent for families and by 24 percent for unrelated individuals. The direction of these results, if not their magnitude, is to be expected, since total income and hence the mean of the combined CPS distribution was raised by 18 percent in the process of bringing each type up to the control total.

Chart I, on the other hand, shows the effect of the inflation procedure on the relative distribution, as illustrated by the corresponding Lorenz curves. Since greater correction ratios must be applied to transfer incomes, a more important source of income for lower income groups, and to property incomes, a more important source for upper groups, than to earnings, the effect—admittedly small—is to reduce inequality at the bottom of the distribution and increase it at the top. The Lorenz curves for the distributions for both families and unrelated individuals therefore intersect, although the effect is more noticeable at the top than at the bottom of the distribution.

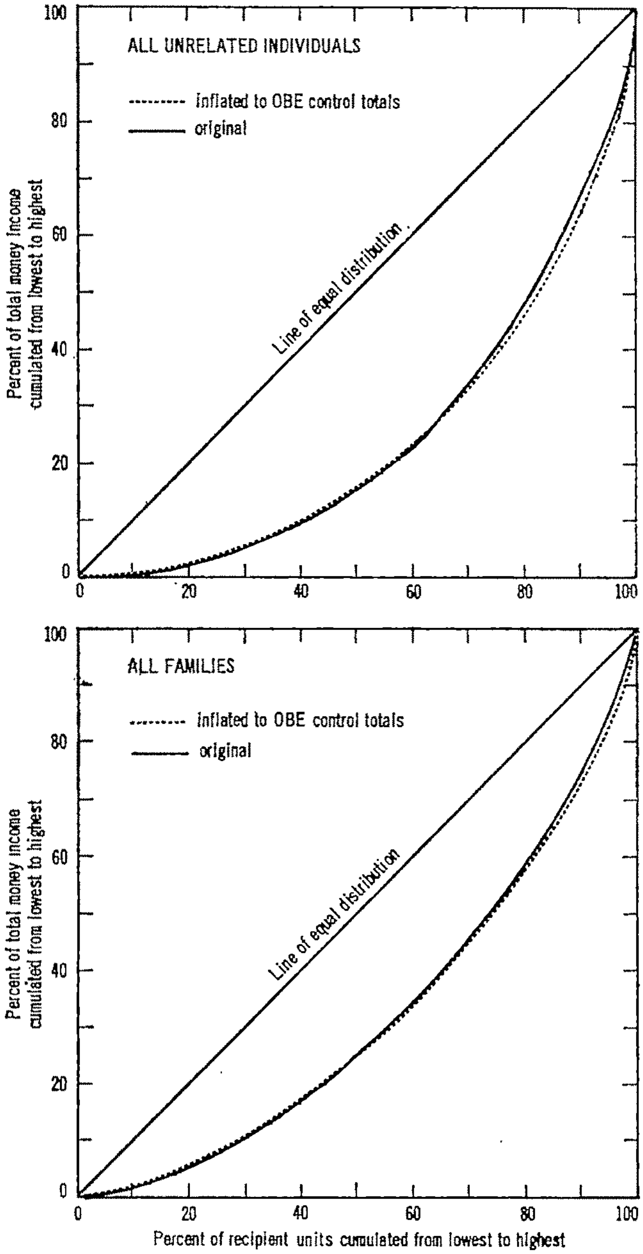
These results are simply designed to be illustrative of the effects of underreporting and should not be interpreted as precise measurements. For one thing, the correction method employed, which, it should be noted, is not necessarily the one to be used in the final series, assumes a constant rate of underreporting of each income type. Even more important, it does not take account of nonreporting; i.e., the failure of many to report certain income types at all. The correction ratios for those who do report the type are consequently overstated.

Other important sources of data on income size distribution are provided by administrative records such as social security records and individual income tax returns. It is well recognized, of course, that tax returns cannot by themselves form the basis for a size distribution: the tax return unit is not a useful income recipient concept, corresponding neither to families, to spending units, nor to individual persons; tax returns do not cover the entire distribution, being restricted to those who file; little socioeconomic information needed in conjunction with an income size distribution is available.

IRS income data are also far from

CHART I

Lorenz Curves for Distribution of Money Income from March, 1965 Current Population Survey



Source: Based on OBE tabulations of the March 1965 CPS, reweighted and nonrespondents allocated by OBE.

ideal. Adjusted gross income (AGI) is clearly not the income concept we want for a size distribution, partly because most transfer income is excluded, partly because the coverage and distribution of certain income types, such as self-employment income, are unduly influenced by tax considerations. Further, income on tax returns is itself not fully reported, judging by its correspondence to control totals. For 1964, for example, while wages and salaries reported to IRS were about 98 percent of the corresponding OBE control, nonfarm self-employment income was 83 percent; property income, 67 percent (with most of the deficiency in interest and rent); and farm income, only 35 percent. Adjusting these estimates for audit results would raise the figures quoted by a dozen percentage points for farm self-employment income, fourteen points for nonfarm, and by just a few points for property income; the wage and salary figure, on the other hand, would be virtually unchanged. Even if not perfect, however, IRS coverage of income, particularly property income, is a definite improvement over field surveys. This improvement is particularly important for the upper income groups; field surveys have rarely been able to capture the upper tail of the distribution so evident in tax data.

While the preparation of a comprehensive income size distribution series, one that accounts for all incomes and income types, both monetary and imputed, must draw on a variety of data sources, particular reliance must be placed on field surveys and tax returns. This was indeed true of the old OBE series, which was described in its original publication as "consisting of the systematic combination of tax returns and survey data to derive comprehensive income size distribution estimates."[1] Although the series was based largely on tax returns, survey data were used for income types not covered by returns and for converting tax return units into family units. This series was not, of course, without its own limitations. The method of estimation, for example, precluded a breakdown by more than the three component distributions: for farm operator families, nonfarm families, and unattached individuals. Heavy reliance had to be placed on bench mark estimates for these three components. Since the studies on which these bench marks were based had become obsolete by the early 1960's, the series was finally discontinued following publication of preliminary estimates for 1963.

In cooperation with the Bureau of the Census, work is now well under way at OBE on reconstruction of this series, utilizing computer techniques for working with individual records and drawing on the basic data sources made available to us on computer tape, such as the CPS, the SFCC, and IRS's Tax Model and audit study. (The Tax Model is a subsample of about 95,000 returns drawn from the Statistics of Income file.) Our ultimate objective is to produce a computer tape with a sample of observations, containing information on the demographic and economic characteristics and the types and amounts of family personal income received by consumer units (families and unrelated individuals) and the persons comprising them. This should permit considerably more flexibility than was possible with the old series in obtaining component distributions defined by demographic variables or by type of income recipient, and in analyzing the determinants, or the consequences, of income distribution through relating income and income types to various characteristics of recipient units.

---

[1] *Income Distribution in the United States, by Size, 1944–1950,* a supplement to the *Survey of Current Business* (Office of Business Economics, U.S. Department of Commerce, 1953), p. 2.

A major purpose of the new series is to relate it more closely to the CPS than was the old. The recipient unit concept used by each was always virtually the same; the new series will simply incorporate the population base provided by the CPS, with the number of consumer units—families and unrelated individuals—taken as of the March survey. While it would probably be desirable to correct an income distribution for changes in family composition and numbers during the income year and between the end of the year and the subsequent March, data for doing so are not currently available. We also propose to establish a closer connection on the income side than existed before, so that a reconciliation of the new series with the CPS series on the distribution of money incomes can be made. Such a relation requires a link between tax return incomes and incomes reported in the CPS; the development of such a link is the single most important step in the new methodology. The ideal method, of course, would be to match the individual recipient units in a field survey directly with the tax returns they filed. Such matching studies were carried out for samples drawn from the decennial Censuses of 1950 and 1960. More recently the Social Security Administration in its pilot link project has been working on a three-way match among a sample drawn from the CPS and their corresponding social security records and tax returns for 1963 and later years. While space is lacking to comment fully on these various link projects, we do not believe at the present time, at least, that they constitute an answer to the problem, partly because of the length of time required to carry them out under methods currently employed, partly because of difficulties in interpreting nonmatches, which cannot simply be equated with nonfilers, and partly because of problems of confidentiality, which make the resulting data files virtually inaccessible to all but a few.

An alternative procedure which we have developed at OBE for linking individual CPS records with individual tax returns is through a statistical rather than an exact match; it may be described as a process for deciding who in the CPS can be expected to have filed a return, and finding a return for them similar to the ones presumably filed, based on characteristics of persons common to both the CPS and the tax return. Unfortunately, the amount of information other than income available on individual returns and the corresponding files such as the Tax Model is quite limited, being largely confined to marital status and the number of taxpayer and dependency exemptions. The first step was to separate returns, using these variables, into groups which are reasonably analogous to groups which can be established from CPS demographic characteristics. Husband-wife couples living together can be taken as analogous to joint and separate returns; other family heads, to heads of households, surviving spouses, and single returns with one or more dependents; other relatives and unrelated individuals (the sum of whom will be referred to as other persons), to single returns with no dependents. The age exemption permits a further classification into under 65 and 65-and-over, giving six groups in all. When due account is taken of probable nonfilers of tax returns among CPS units, the number of CPS units and of tax returns in each group in 1964 appeared to be sufficiently close to make the classification a reasonable one.

Within each group, wage and salary income was selected as the principal linking variable, because it is both relatively well reported in each source and widely received. As might be expected from evidence cited earlier, within each group the CPS distribution of wage incomes was generally lower than the corresponding tax return distribution. When account was taken of the difference in the means of the two distributions and the presence of nonfilers in the CPS, however, it appeared that there was sufficient resemblance between the distributions to warrant use of this link.

It would be unfair to gloss over all the difficulties faced in this step. The correspondence between the groups is not exact: joint returns, for example, may be filed by married couples not living together; a person may qualify as head of a household if he is living alone but supporting, say, his aged parents living elsewhere. In the absence of outside information on the types of returns filed by persons of different family status, some limited rearrangement of returns assigned to each CPS group from that indicated above was necessary in order to improve both the estimates of nonfilers in each group and the correspondence between the two wage distributions.

Space is lacking to give more than a summary description of the methods by which the match was carried out and the variables on which it was based, much less to justify all the methods chosen. Since in all groups the number of CPS units exceeded the number of returns, it was necessary to determine who in the CPS was not to be assigned a return. Weight was given to the existence and size of self-employment income as well as to taxable income, since filing requirements for the former are lower than for adjusted gross income generally. CPS units excluded from the match all had taxable incomes[2] below the filing requirements for adjusted gross income. After this step, the number of units in each group in the two distributions was the same.

Next, within each group CPS units and tax returns were ranked by size of wage income. The recipient units were then separated into a specified number of wage rank classes containing an equal number of units in each source; e.g., the top one million units in any group. The number of classes had to be large enough so that there would not be too much disparity in CPS and IRS wage incomes within the class and yet small enough to give some weight to the other income types used in the match. In selecting the breaking points for a class, particular care was used to avoid round-off points in the CPS; e.g., incomes in multiples of $100 or $500 where observations tend to cluster.

Within each of the wage classes, units with nonfarm self-employment income in each of the two sources—CPS and IRS—were ranked by size of such income and then matched according to their rank; similarly for farm income, and for records with both self-employment types. The remaining records, including those with no self-employment and those with self-employment where there was an excess of such records in one source over those in the other (e.g., an excess of records with farm income in the CPS over those with Schedule F income in the Tax Model), were ranked by size of property income and matched by their rank. For records without such income, the ranking was on a random basis. The net result, for this last group, was that CPS units reporting property income nearly always received a tax return with a larger property income than the CPS amount. Further, since in each class the number with such income in IRS nearly always exceeded the number in the CPS, many CPS units who reported none in CPS were assigned, on a random basis within the class, a tax return with property income. In all cases, a "none" in CPS was given a tax return amount no greater than that received by a CPS unit reporting a positive amount.

---

[2] Taxable income was defined as earnings plus property income plus the "all other" category in the CPS. Four separate income types other than earnings were reported and coded separately in the CPS for 1964: (1) social security and railroad retirement; (2) property income; (3) unemployment compensation and public assistance; (4) all other, including pensions, annuities, veterans' payments, workmen's compensation, royalties, and interpersonal transfers. Perhaps half of "all other" income is subject to income tax.

It might be useful to outline a few consequences of these methods for the earnings types. For those CPS units with wage income, almost none received a tax return reporting no wage income, and for a substantial majority, although certainly not all, the tax return wage income exceeded that reported in the CPS. There was no guarantee, although a high probability, under our rules, that a CPS unit reporting self-employment income would be matched with a return with that same income type. Indeed, 90 percent of those with nonfarm in CPS received a return with business or partnership income and 78 percent of those with farm received a return with Schedule F income. If we exclude from the calculation about a half-million CPS units reporting farm income who could not be given a farm return because of the deficient number of the latter, the matching rate would be 90 percent. These percentages, incidentally, are similar to the percentage of those in exact matches who are found to have reported a given self-employment type in both sources.

Space is lacking for a full presentation of the reasons for selecting these rules rather than others, or for speculating on the difference alternative rules would have made to the results. Cross-tabulations between wage and property income for both the CPS and the Tax Model confirmed the relationship between the existence and size of these two income types in both sources; for the matching procedure, the basic assumption required is that those reporting property income in the CPS are more likely to have reported it on their tax return and to have reported a larger amount than those who reported none.

It was decided to give primacy to wage income over self-employment income as a link, by matching self-employment records within each wage class, largely because the former is the more consistently and the more accurately reported of the two types. This is suggested by matching studies and by the closer relationship between CPS and IRS wage distributions than the corresponding self-employment distributions. For nonfarm self-employment, the tax return distributions show considerably more dispersion than the CPS, with more top income recipients, and many more with low, and particularly negative, incomes. The reasons for these disparities in the self-employment distributions are not fully understood. It has been alleged that respondents tend to report gross rather than net income in the CPS, although there is little direct evidence to support this contention. Perhaps there is also some tendency to report "normal" self-employment income in the CPS rather than last year's actual income. With respect to the tax data, a greater degree of underreporting in IRS at the bottom of the distribution than at the top seems to be indicated by the audit study. Part of the difference may be due to more liberal tax treatment of expenses, such as depreciation, which the self-employed may really consider as part of their net income.

While these are reasons for emphasizing the wage link, they are not sufficient for discarding self-employment entirely. As noted, the evidence from matching studies suggests a high degree of consistency in reporting its existence, as well as indicating some direction relationship in the amount reported. Further, self-employment is particularly important in the class with no wage income: about half of those with self-employment have no other earnings type, and a high proportion of filers with no wage income have self-employment; 80 percent of the joint returns with no wage income, for example, had business, farm, or partnership income. Cross-tabulations between wage income and self-employment income also indicate a similar pattern in both sources. The matching of self-employment records should therefore be improved by first taking wage income into account.

The statistical match we have carried out can serve several different interests. One might be tax research based on the Tax Model, where there is a need to use consumer units rather than tax returns as the population base and where some estimates of nonfilers and their corresponding incomes (drawn from the CPS) are required. The Brookings Institution has experimented with converting the 1962 Tax Model from tax returns to consumer units but without the population base provided by a field survey. As a result, it has been difficult to devise a method which would incorporate nonfilers and their incomes into the Brookings tax file. Our file might also be usable in work on income maintenance programs and the negative income tax, where data for consumer units on income other than AGI and its components are required.

As already noted, the major purpose of the match for our project is to correct and amend the CPS income figures. As a preliminary exercise, it is worthwhile to see what kind of difference in the size distribution of total money income and in the measurement of poverty would result from substituting, for those units we defined as filers, the tax return amount for the CPS amount for those income types which the two sources have in common: wages and salaries, farm and nonfarm self-employment income, and property income. The results, before and after carrying out these substitutions, are shown in Tables 3 and 4, separately for families and unrelated individuals.

Before commenting on the results, note

TABLE 3

DISTRIBUTION OF MONEY INCOME IN 1964 FOR CONSUMER UNITS, BASED ON THE CPS AND ON TAX RETURN DATA
(Millions of units)

Total Money Income	Families*		Unrelated Individuals	
	CPS	Adjusted by Tax Returns	CPS	Adjusted by Tax Returns
Under $ 1,250.........	2.83	2.68	4.32	3.82
$ 1,250- 2,749.........	5.83	5.34	2.94	2.97
2,750- 4,749.........	8.34	8.16	2.19	2.40
4,750- 6,749.........	9.69	8.93	1.46	1.51
6,750- 9,749.........	11.42	11.84	.75	.94
9,750- 14,749.........	7.98	8.50	.25	.28
14,750- 24,749.........	2.47	2.80	.09	.09
24,750 and over.........	.55	.87	.04	.04

* The sum of families and subfamilies as defined by Census.

"CPS": Same as "Original" in Table 1.

"Adjusted by tax returns": CPS after substituting, for filers, the tax return amount for the CPS amount of wages and salaries, farm and nonfarm self-employment income, and property income.

SOURCE: Office of Business Economics.

TABLE 4

CONSUMER UNITS ABOVE AND BELOW POVERTY IN 1964, BASED ON THE CPS AND ON TAX RETURN DATA
(Millions of units)

	Families*		Unrelated Individuals	
	CPS	Adjusted by Tax Returns	CPS	Adjusted by Tax Returns
In poverty	7.60	7.33	5.04	4.61
Not in poverty	41.52	41.79	7.01	7.44

* The sum of families and subfamilies as defined by Census.
Definitions and sources: Same as Table 3. Poverty line as defined in Table 2.

should be taken of the following: First, since the population universe for the comparisons is all consumer units, nonfilers have been included in both distributions with their CPS incomes. Second, nontaxable income types for filers, in particular, social security, unemployment compensation, and public assistance, have been included in the IRS distribution. Third, the "all other" income category in the CPS has been used in both distributions, since it is not possible to make this component comparable with the miscellaneous category in the Tax Model. While both sources include royalties, the coverage of pensions and annuities differs; the CPS includes such nontaxable income types as veterans' payments and workmen's compensation; and IRS contains items, e.g., gambling gains, net operating loss deduction, which have no counterpart in the CPS. Fourth, the IRS farm figures have been used for the comparisons, even though their deficiencies make it unlikely they will form the basis for the farm estimates in the final OBE series. None of these comments, it should be noted, bears on the CPS distribution; it is therefore the same as the original shown in Tables 1 and 2.

In Table 3, the effect of the tax return substitution for families is to reduce frequencies moderately in the lower income brackets and raise them in the upper ones; for the two top brackets combined, the number of families is increased by 20 percent. For unrelated individuals, the reduction is confined entirely to the bottom bracket, with no observable effect on the upper tail. As might be expected, the impact on poverty levels in Table 4 is minimal; the number of consumer units in poverty is reduced by some 700,000, or by less than 6 percent of the original CPS number.

Part of the difference between these two distributions arises from the fact that the income aggregate implicit in that for tax returns is 7 percent more than the CPS; indeed, substituting tax return earnings and property income into the CPS distribution serves to cut the underreporting rate of 15 percent previously referred to for the CPS to 9 percent. It is interesting to see how much of the difference between the two distributions and the two estimates of poverty remains when the two sources are adjusted to reflect the same mean income. For this purpose, the "adjusted" CPS estimates given in Tables 1 and 2, which have already been brought up to the OBE control for Census money income, can be compared with the tax return distribution where the latter's earnings and property income have been inflated to the same control as that used for the Census types by the same method; multiplying each type by the ratio of the

TABLE 5

DISTRIBUTION OF MONEY INCOME IN 1964 FOR CONSUMER UNITS, BASED ON THE CPS AND ON TAX RETURN DATA, ADJUSTED TO OBE MONEY INCOME CONTROL TOTAL
(Millions of units)

Total Money Income	Families*		Unrelated Individuals	
	CPS	Adjusted by Tax Returns	CPS	Adjusted by Tax Returns
Under $ 1,250........	2.01	1.99	3.03	2.82
$ 1,250– 2,749........	4.33	4.35	3.34	3.28
2,750– 4,749........	7.55	7.84	2.24	2.46
4,750– 6,749........	8.92	9.25	1.62	1.70
6,750– 9,749........	11.48	12.07	1.09	1.12
9,750– 14,749........	10.09	9.15	.47	.44
14,750– 24,749........	3.79	3.32	.18	.17
24,750 and over........	.96	1.17	.09	.07

* The sum of families and subfamilies as defined by Census.
"CPS": Same as "Adjusted" column in Table 1.
"Adjusted by tax returns": Same as "Adjusted" column in Table 3, except that each income type was inflated to the OBE control for money income for that type.
SOURCE: Office of Business Economics.

control to the actual amount in the Tax Model, separately for families and unrelated individuals. The results are shown in Tables 5 and 6, and the Lorenz curves for the two money income distributions are portrayed in Chart II.

In view of the adjustments made, it is, of course, impossible for the frequencies in the lower brackets in one distribution to be uniformly smaller than those in the other, and in the upper brackets, uniformly larger. For families, frequencies for the tax return distribution are virtually unchanged for the two bottom brackets, increased by some 4 percent in the brackets from $2,750 to $9,750, reduced by 11 percent in the brackets from $9,750 to $24,750, and increased by 22 percent in the top bracket, $24,750 and over. In addition, it is interesting to note that the mean income in the open end bracket for the tax return distribution exceeds that in the blown-up CPS by more than $3,000. This capture of the upper tail of

TABLE 6

CONSUMER UNITS ABOVE AND BELOW POVERTY IN 1964, BASED ON THE CPS AND ON TAX RETURN DATA, ADJUSTED TO OBE MONEY INCOME CONTROL TOTAL
(Millions of units)

	Families*		Unrelated Individuals	
	CPS	Adjusted by Tax Returns	CPS	Adjusted by Tax Returns
In poverty..............	5.52	5.76	3.83	3.65
Not in poverty..........	43.60	43.37	8.22	8.40

* The sum of families and subfamilies as defined by Census.
Definitions and sources: Same as Table 5. Poverty line as defined in Table 2.

the family distribution from tax return data is also evident in Chart II. The Lorenz curve for tax returns lies below that for the CPS throughout. The two percentage point gain in the income share of the top 5 percent is largely at the expense of the 50th through the 95th percentiles rather than of the bottom half of the distribution. If we compare the blown-up tax return distribution with the original CPS shown in Chart I, the difference between them is even more striking: the two curves intersect just below the 50th percentile, with the 3 percentage point gain of the top 5 percent being entirely at the expense of those from the 50th to the 95th percentile.

While space is lacking to comment at any length on the results for unrelated individuals, the effect of substituting tax return data seems to be quite the opposite of that for families: frequencies at the two extremes are reduced and those in the middle are increased, with the tax return Lorenz curve in Chart II lying above that for the adjusted CPS. The difference between the two distributions for the measurement of those below poverty is again minimal, although in the direction that would be expected from the already observed changes in the distributions for families and for unrelated individuals: a minor increase in the number of poor families and a minor reduction in the number of poor unrelated individuals.

The tax return distribution shown above should not be interpreted as necessarily resembling the final OBE distribution of money income (Census concept). Even though the former does come up to OBE's control totals for money income, the inflation procedure employed does not take account of additional information which we anticipate using in going from tax return and CPS income types to their corresponding controls. We are currently working on incorporating the results of the 1963 audit study into the tax return data used in the match; this step alone should reduce the difference between the actual tax return distribution and the control by two percentage points. Data from the SFCC can be used for distributing accrued interest on savings bonds and interest on state and local securities and perhaps for increasing the proportion reporting interest and dividends. Information from special studies conducted by HEW on characteristics of persons receiving particular kinds of public assistance payments can be used to improve the CPS coverage of this income type. If available in time, the SEO can be used for correcting certain other transfer incomes. Use of these and other sources should reduce, although they will not eliminate entirely, the need for some type of inflation technique to bring the final results up to control totals.[8]

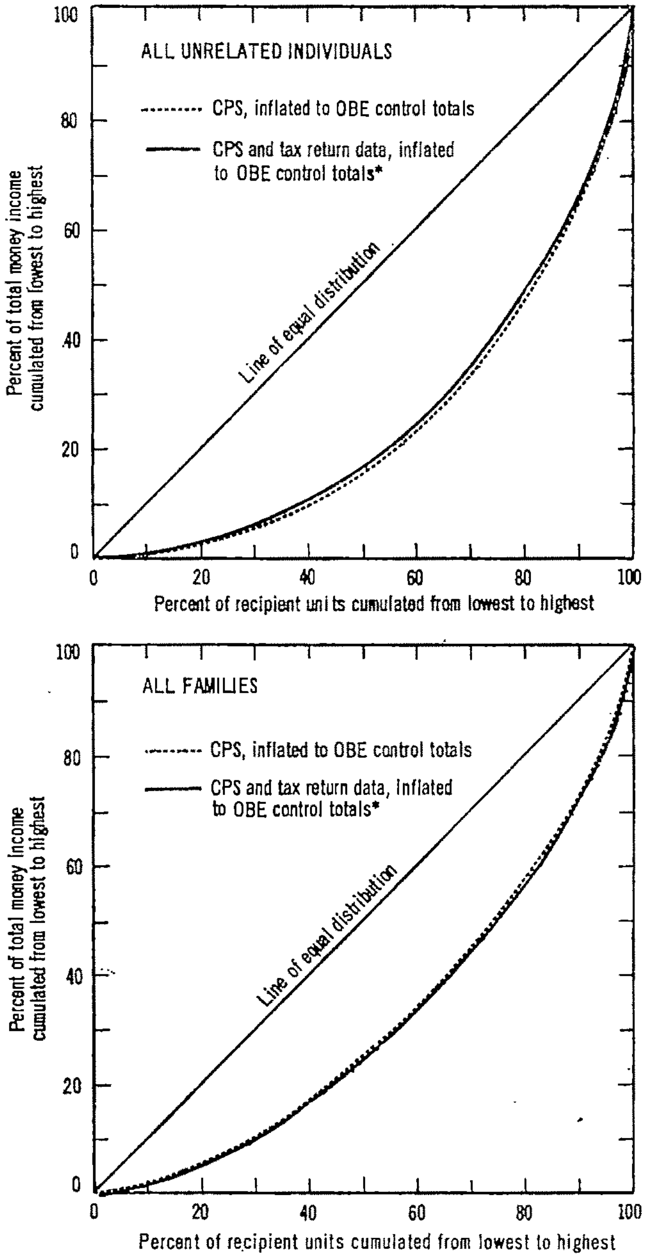
Nothing has been said about the inclusion of nonmoney income in the final distribution. The estimation of such imputed income types as rent on owner-occupied dwellings and imputed interest will probably have to be based on asset data from the SFCC or the SEO. Space is lacking for a discussion of methods to be used for these income types, and results of our work, other than the few we have given for the match, are not at present available. Complete results for 1964 should be forthcoming some time next year. Estimates for subsequent years will be developed by means of the techniques used for 1964, but it is impossible to specify their timing at this moment. Since the data sources we are using are not available in

---

[8] Such a technique need not, of course, be of a constant multiplicative type. In any case, it is obvious that the same multiplier cannot be applied to both positive and negative incomes. In the inflation technique used in this paper, the reciprocal of the multiplier for positive incomes was used for loss incomes.

CHART II

LORENZ CURVES FOR DISTRIBUTION OF MONEY INCOME, BASED ON THE MARCH 1965 CPS AND ON TAX RETURN DATA FOR 1964



*CPS after substituting, for filers, the tax return amount for the CPS amount of wages and salaries, self-employment and property income.

Source: U.S. Department of Commerce, Office of Business Economics the same form for years prior to 1964, there are no plans to estimate a size distribution for earlier years by the new methods. The suitability of some extrapolation device to provide a link with the earlier OBE series remains to be explored. Even if a precise link to the past cannot be provided, the new series should be important in determining the current size distribution and estimates of those below poverty levels and in providing data that should be useful in understanding the sources and consequences of inequality prevailing in a given year.

MEASURES OF ECONOMIC WELL-OFFNESS AND THEIR CORRELATES

By JAMES N. MORGAN, *University of Michigan*
and
JAMES D. SMITH, *Pennsylvania State University*

We are reporting some early findings from the first wave of a panel study of the dynamics of economic well-being. Since we are just completing the organization of the first wave of survey data we can say little about dynamics. Those comments must await at least a second interview of the panel. However, it is worthwhile to present some static correlations from the first interview which have meaning for the way one might choose to measure how well off families are.

One of the problems of measuring well-offness comes about because consumers live in families and families differ, not only in the amount of income they receive, but also in composition. It is obvious that x dollars of income implies one level of well-offness for a two-person family and quite a different level for a six-person family. What is needed is some adjustment for the economic size of a family. The simplest way is to divide family income by the number of family members. This per capita measure of income is certainly an improvement over a per family measure. Yet it can be argued that the "needs" of an adult are quite different than those of a child and to give them equal weight in a per capita measure of income may fall short of a measure of well-offness which is neutral with respect to family composition.

It is possible to go further and estimate normative quantities of goods and services needed by each type person, convert these to market values, and sum them to measures of family need. In this way we can account for differing needs associated with age and sex. It is normative in the sense that differential needs by age and sex have been set on the basis of some noneconomic judgment. The absolute quantity of goods for any type person is not important; it is the differentials among different types of people which must be consistent.

There is a long history of attempts to adjust measures of family welfare to allow for the different sizes and shapes of families, and hence their different levels of need. Even in the United States, it started two centuries ago with estimates of food needs; then was elaborated to clothing and housing and finally to total needs.[1]

Recently, economists have gotten into the act, and although they have not invented picturesque names like "ammain," adult main maintenance scale (Sydenstricker and King) or "quet" (Engel in honor of Quetelet), they have suggested —and in a few cases carried out—ingenious econometric methods for estimating the components of equivalent adult scales from behavioral data rather than from normative standards.[2]

S. J. Prais actually estimated jointly (by iteration) the income elasticities and the coefficients for family members of each age-sex, for various food components, and total food. The coefficients are remarkably close to the National Research Council's calorie requirements, and to the NRC protein requirements, and to the food standards currently used by the U.S. Department of Agriculture which form the basis of the Orshansky-type poverty thresholds as well as Family Equivalence scales of the BLS.[3]

---

[1] An excellent survey of this development and a bibliography of studies of family living can be found in Faith M. Williams and Carle C. Zimmerman, *Studies of Family Living in the United States and Other Countries: An Analysis of Material and Method,* U.S. Dept. of Agric., Misc. Pub. No. 223 (U.S. G.P.O., Washington, D.C., Dec., 1935).

[2] See for instance Milton Friedman, "A Method of Comparing Incomes of Families Differing in Composition," in *Studies in Income and Wealth* (N.B.E.R., 1952) and S. J. Prais, "The Estimation of Equivalent Adult Scales from Family Budgets," *Econ. J.*, Dec., 1953, pp. 791–801, and J. A. C. Brown, "The Consumption of Food in Relation to Household Composition and Income," *Econometrica,* Oct., 1954, pp. 444–60.

In fact, the major difference is that the USDA food requirements explicitly allow for economies of scale in feeding people, after adding up the individual requirements.

It is these USDA standards which form the basis for the poverty thresholds developed by Mollie Orshansky, of the Department of Health, Education, and Welfare. Additional economies of scale are allowed for by using three times the food requirement as a total needs estimate except for families of one or two, where a larger multiplier is applied. And an adjustment is made for nonfarm families to allow for nonmoney incomes. At one time it was argued that farm families could get along on 60 percent of the income of a city family. The official Orshansky figures use 70 percent and a recent publication of the Department of Agriculture suggests that it takes a farm family at least 85 percent as much.[4] And of course there are several different levels of adequacy for which the USDA and the HEW standards are published.

Friedman[5] in a paper written in 1935 but published in 1949 suggested an imaginative method of fixing the equivalence of income among families of various size and composition. The method calls for first fitting a cross-sectional consumption function for a size and composition of family which will serve as "standard," say, a husband-wife family. The estimated values of the constants from the function, A and B in the linear form:

$$T = A + Br,$$

where T stands for total consumption of food and r for family income, are then applied to other types families with the condition that:

$$\frac{T_{ij}}{S} = A + B\frac{r_{ij}}{S}$$

where S is an equivalency value and T_{ij} and r_{ij} are the total food consumption and income of the ith family of the jth type. Since the values of A and B are known from the function fitted to the data for the standard family, and T_{ij} and r_{ij} are known for each other family, the value of S can be taken as:

$$S = (T_{ij} - Br_{ij})/A$$

The next step is to evaluate the presumed determinants of S: age, sex, and number of family members, with account taken of the marginal contribution made by each additional family member. This can be done by taking 1 as the value of the standard husband-wife unit and estimating separately the increment to S of each additional member of each sex.

$$S_{nk} = 1 + a + b_1L_1 + b_2L_2 + \cdots b_mL_m$$

where S_{nk} is the value of S with the addition of the nth family member of the kth sex and L_1, L_2, . . . L_m are dummy variable for a suitable set of age intervals.

There have been a few applications of

---

[3] *Recommended Dietary Allowances*, 6th Rev. Ed., NRC Pub. 1146 (National Academy of Sciences-National Research Council, 1964), and "Cost of Food at Home," *Family Econ. Rev.*, periodically, e.g., Mar., 1967, p. 20, and "Estimating Equivalent Incomes or Budget Costs by Family Type," *Monthly Labor Rev.*, Nov., 1960, pp. 1197–1200.

[4] Published talk by Carol M. Jaeger to the Consumer and Food Economics Research Division, 45th Annual Agricultural Outlook Conference (Washington, D.C., Nov., 1967).

[5] *Op. cit*

the Friedman technique. On a limited basis, Jean Mann Due[6] used a sample of families in Portland and data from the purchases study to estimate an index of equivalent family income. She fitted basic regression lines for total expenditure and food expenditures, separately. Harold W. Watts[7] in an exploratory article applied the approach to grouped data from the 1960 Survey of Consumers Expenditures.

Due discovered that age of family members was uncorrelated with either food or total family expenditures; family size was the only important dependent variable in the determination of the equivalency ratio. Watts proceeded by fitting separate multiple regressions of food expenditures and "necessities" expenditures on income, family size, region of the country, urban-rural (nonfarm) residence, age of family head, and homeowner status. However, no evidence was presented that he tested the importance of the age of family members, other than the head, on the consumption of food or necessities. He does, however, report important evidence on the impact of geographic location, on the equivalency ratios. Looking at his most dramatic locational difference, it takes, on the average, for all sizes of families and all levels of income about 68 percent more income in the urban Northeast before the proportion of family income spent on food is equal to that in the rural South. If one takes some account statistically of some of the weaknesses of the original approach—for instance, the assumption that equal expenditure proportions for food mean equal utility across family types and income levels, as Watts does—the locational differences are compressed but remain important. These assumptions are rather critical because what the Friedman-type approach offers is objectivity and pure economic measures. If one can alter the equivalency scale by arbitrarily choosing among consumption categories, and there are no powerful, if not conclusive, arguments for selecting a specific category of consumption, the value of the exercise is questionable. If, for instance, one were to choose clothing expenditures over income as a criterion for equal levels of well-offness, and as a consequence the ranking of families changed from what it had been with levels of well-offness deemed equal when the proportion of income spent on food was equal, then the usefulness of the scale is dubious.

A less serious, but by no means inconsequential flaw, would be the case in which the rank order of families was unchanged, but the index of well-offness changed because there is not a fixed relation between the ratio of food expenditures to income and clothing expenditures to income across family types. The implications are clear; one must at least construct a weighted index of consumption categories over income if he is going to avoid this problem. Friedman recognized that it might be necessary to weight consumption expenditures, but rejected the procedure on pragmatic grounds: ". . . its adoption would render the statistical procedures too complicated."[8]

Empirical Exploration

In the remainder of this paper we will present some empirical results from the manipulation of unweighted data from a panel study of economic dynamics we are doing under a grant from the Office of Economic Opportunity. The study calls for following all the members of about

---

[6] See appendix to Friedman, *op. cit.*

[7] Harold Watts, "The Iso-prop Index: An Approach to the Determination of Differential Poverty Income Thresholds," *J. of Human Resources*, Winter, 1967, pp. 1–18.

[8] *Op. cit.*, p. 19.

5,000 families (about 20,000 individuals) over the next five or six years of their lives. The basic sample design includes about 2,000 low-income families taken as a subsample of an unpublished survey of 30,000 families interviewed by the Census Bureau for OEO in 1967 and a national cross-section of about 3,000 families drawn from the Survey Research Center sampling frame.

The OEO subsample was specifically designed to overrepresent the lower tail of the income distribution and only families with incomes of less than \$2,000 + \$1,000 N ($N =$ number of persons in family) in the OEO sample were eligible for selection.[9] We will refer to this as the "poor" sample. The other part of the sample was drawn from the Survey Research Center sampling frame. It is a self-weighting sample, except some minor adjustments will be made for differential nonresponse in the Northeast urban areas. We will refer to this as the "cross-section." Sometimes we have used both samples together and refer to it as the "combined" or "whole" sample.

It is our position that a reasonably good ranking of individuals can be constructed from ratios of family income to a normative needs standard. We found that such a ranking was relatively insensitive to changes in the types of adjustments Orshansky made for farmers and for small families and for the level of the food standard.

What may matter, however, are better measures of income and broader measures of well-offness which include measures of leisure. Some time ago, Martin David suggested including nonmoney incomes and assets in a ratio to needs.[10]

We have constructed a real income measure which includes imputed rent, returns from do-it-yourself activities, and food, clothing, and housing received from employment. We have also deducted some costs of getting an income: federal income taxes, union dues, and child-care costs when all adult family members are working. We propose looking at the ratio of this income measure to a needs standard (which takes account of family size and farm status) in conjunction with leisure time left to enjoy income, quality of housing and neighborhood, and the ratio of income to the cost of the USDA food standard.

Having gone from a dollar income to a ratio of income to needs, what can we do about the important dimension of leisure time? One of us proposed last year at these meetings that a reasonable procedure would be to multiply the leisure hours by the ratio of income to needs, on the assumption that leisure and income are substitutable commodities in a preference map of rectangular hyperbolas. The result of this multiplication is a pure number which can be used for ranking persons by well-offness.

One can argue that some people love their work and others hate their spare time, but we can argue that at least at the margin in a society like ours where people have some choice about overtime, extra work, working wives, etc., the values of additional leisure and money are equatable. But how does one measure leisure? If one deducts ten or twelve hours a day for sleeping and eating, some with two jobs will have zero or negative leisure and hence zero or negative well-offness. By allowing only eight hours a day for sleeping and because there are weekends, we have avoided this. We also counted only the head and wife, not the children, whose leisure is difficult to evaluate. And we subtracted from leisure sixteen hours a day during illness, and eight hours a day during unemployment, as hardly "enjoyable leisure," and an estimate of the hours required for housework and child care, based on data from a previous study.[11]

---

[9] Eligibility was also limited to OEO families whose head was under age sixty. The original OEO sample was itself a complex probability sample and any use of the sample for making national estimates, as opposed to examining relation between variables, must await the arrival of detailed sampling information not available to us at the time of this writing.

[10] Martin David, "Welfare Income and Budget Needs," *Rev. of Econ. and Statis.*, Nov., 1959, pp. 393–99. This was done extensively in *Income and Welfare in the United States*, Morgan, David, Cohen and Brazer (McGraw-Hill, 1962).

Figure 1 gives some notion of the order of magnitude of the adjustments we are talking about. For a representative cross-section of the population the average family money income in 1968 was $8,400. If we calculate for each family its basic dollar needs *à la* Orshansky, it comes to $3,369 for the average family. And if we calculate the ratio of income/Orshansky for each family and average these ratios, the average ratio comes to 2.6. (Note that an average ratio is not the same thing as the ratio of averages.)

Conservative estimates, obtained by asking people about nonmoney components of income, come to $695 for the average family. This does not include capital gains but does include estimates of the value of housing received by owning a home or living with relatives or getting housing as part of a job. It includes estimates of the value of do-it-yourself projects such as fixing the house or the car, growing food, and sewing. Also included are food and clothing received free or as part of one's pay or through food stamps.

And we subtract estimated federal income taxes, union dues, and payments for child care when all the adults are working, which come to an average of $950 per family. There are many other things that could be considered costs of earning income, but they were either likely to be a relatively stable fraction of income or were impossible to measure. We may try this year to add one more—the cost of the journey to work—though it may well affect leisure time more than it affects net money income. Anyway, the cost of getting income was $950 for the average family.

As you can see in the lower half of Figure 1, if we improve the measure of control over resources, divide by a needs standard without adjusting its scale, and multiply by our measured leisure, we have an index of well-offness which comes to 8300 for the average family. (Again this is a mean of the measure, not the result of dividing and multiplying means.) The figure gives some correlations, indicating that family size and hence needs are positively correlated with income but nega-

1. "Orshansky" ratio
   $a/b=c$
2. Well-offness ratio

$$\frac{de}{b}=f$$

$r_{ab}=.33$
$r_{ac}=.85$
$r_{db}=.33$
$r_{de}=-.37$
$r_{be}=-.42$
$r_{d/b,e}=-.19$
$r_{df}=.20$
$r_{bf}=-.24$
$r_{ef}=.10$

Where:

a = Family money income
b = Needs standard
c = "Orshansky" ratio
$d=a+$nonmoney income$-$tax and other costs of getting income
e = Leisure
f = Well-offness
$\bar{a}=\$8,400$
$\bar{b}=\$3,369$
$\bar{c}=2.6$
$\bar{e}=3427$ hours
$\bar{f}=8300$ points

Mean nonmoney income = $695
Mean tax and other costs of getting income = $950

FIGURE 1
COMPOSITON OF TWO MEASURES OF ECONOMIC STATUS

---

[11] Morgan, Sirageldin, and Baerwaldt, *Productive Americans* (Ann Arbor, Inst. for Soc. Res., 1966).

tively correlated with leisure time (because large families require more housework and child care). Leisure time is also negatively correlated with income, largely because a working wife still does most of the housework, too, but also because of substantial variations in the work hours of the family head. As a result of its negative correlations with both the numerator (income) and denominator (needs), leisure ends up with a small negative correlation with the income/needs ratio, and a still smaller positive correlation with the well-offness index.

Of course, the well-offness measure is completely determined by its three components, and in log form is an additive relation. The relative importance of the three components could, therefore, be estimated by calculating the ratios of the standard deviations of their logs, if one were interested in their relative importance in explaining the log of well-offness. Actually the relative variance of the leisure and needs measures are less than two-thirds as large as the relative variance of net real income. What all this comes to is that a more complete measure of economic well-offness is still heavily dominated by real income and real income is dominated by money income. But this is no reason for complacence. For one thing there are differences, and, more important, the domination will almost certainly turn out to be less at the bottom of the income distribution.

We welcome suggestions as to better functional forms and intend to experiment ourselves with variants. But again, we feel that there are more important problems in the measurement of how well off a family is.

A major problem in assessing a family's economic status and its change over time is housing. Determining housing needs (and costs) has long been a major problem for welfare administrators and for those attempting to measure the cost of living. Housing accounts for much of the difference between geographic areas in cost-of-living indexes, and one suspects that much of that difference is spurious—reflecting different packages of social goods that go with the housing and different housing standards. In dealing with our problem of measuring the status of individual families, we reduced the errors by adding to real income the value of free rent received by living with relatives, getting housing as part of pay, or owning one's own home. But some people live in housing projects or at historic low rents left there by inertia or at high rents when they move into new areas.

We have, therefore, developed another measure by subtracting from income the family's outlays for housing, even if they include repayment of mortgage principal. We did not use an Orshansky-type needs measure as a divisor, since it builds in an economy-of-scale adjustment for housing for one- and two-person families; so we used the USDA food needs standard, which is about a third as large as the Orshansky total needs standard. Again we can multiply that ratio by leisure to get a well-offness index. Of course, this then requires some separate assessment of the quality of the family's housing and environment. Then if a family moves from an expensive slum apartment to living free with relatives and then to a subsidized housing project, we can see how their income net of housing goes up, then down a little, and how their housing conditions improve a little, and then a great deal.

We have also developed an index of the quality of housing and neighborhood—again somewhat arbitrarily—by adding a point for each of the following:

1. Enough rooms for the family (assuming two children of like sex could share a bedroom).

2. Value per room at least $2,000.
3. Do not share home with unrelated people.
4. Either have a car or good public transportation.
5. Own their own home.
6. Live in a single family house.
7. In a neighborhood of single family houses.
8. No visible housing lacks reported by interviewer.
9. Live near but not in the center of a metropolitan area of 50,000 or more (5–15 miles from center).

We see no easy way to combine this measure of housing quality with the well-offness measure excluding housing. There is no particular reason to believe that the indifference curve between those two is a rectangular hyperbola. The only reason for multiplying in this case would be that items measured in different units can only be combined by either dividing or multiplying, not by adding or subtracting.

We also asked enough questions to develop an estimate of the family's food consumption, and given the heavy reliance of the needs standards on food needs, thought it sensible to compute a measure of dietary status by taking for each family the ratio of food consumption to its estimated food requirements. The average family spends 1.78 times the standard.

Figure 2 provides a picture of the intercorrelations among all these various measures of family economic status, for a combined sample consisting of a representative cross-section of some 2,900 plus a low-income sample of nearly 1,900 cases.

Keeping in mind that high correlations do not mean that different measures would not change the count of the poor but only that they would have little effect on the ranking of individual families, we notice that they become progressively smaller as our measures become more complex, taking account first of differing family structures, then of differences in

	Income Measures			Ratios = $\frac{\text{Income}}{\text{Needs}}$			$\frac{\text{Income}}{\text{Needs}}$ x Leisure			Other Indicators	
	1. $ Income	2. Net Real	3. Net Real-Housing	4. $\frac{\text{\$Inc}}{\text{Needs}}$	5. $\frac{\text{Net Real}}{\text{Needs}}$	6. $\frac{\text{Net Real-H}}{\text{Needs}}$	7.	8. (See stubs at left side)	9.	10. $\frac{\text{Food}}{\text{Food Needs}}$	11. Housing Quality
1. Money Income											
2. Net Real Income	.99										
3. Net Real Income – Housing Outlays	.99	.99									
4. $ Income/Needs	.85	.83	.83								
5. Net Real Income/Needs	.84	.83	.82	.98							
6. $\frac{\text{Net Real Income-Housing}}{\text{Food Needs}}$	.74	.72	.73	.96	.96						
7. $\frac{\text{\$ Income}}{\text{Needs}}$ x Leisure	.77	.75	.75	.96	.95	.93					
8. $\frac{\text{Net Real Income}}{\text{Needs}}$ x Leisure	.72	.73	.72	.91	.94	.91	.98				
9. $\frac{\text{Net Real Inc.-Hsng}}{\text{Food Needs}}$ x Leisure	.65	.64	.64	.89	.91	.95	.96	.96			
10. $\frac{\text{Food Consumption}}{\text{Food Needs}}$	.35	.34	.32	.52	.54	.54	.52	.53	.53		
11. Index of Quality of Housing and Neighborhood	.42	.47	.43	.39	.45	.34	.38	.43	.33	.23	

FIGURE 2
CORRELATIONS AMONG MEASURES OF ECONOMIC WELL-OFFNESS
(Combined unweighted samples — $n = 4797$)

leisure, and finally in differences in what the family must pay for housing. One indication of the validity of the adjustment for family size is that the food adequacy measure, the next to bottom row, correlates more highly with the ratios of income to needs than it does with any of the absolute income measures.

Interestingly enough, the index of the quality of housing and neighborhood correlates poorly with all the measures and poorest with our best measure of well-offness excluding housing. Other studies have shown inertia in the housing market, and it is reflected here.

These correlations were very much the same when we looked only at the representative cross-section part of the sample, indicating that the differences are not substantially greater among lower-income people.

One final comment on these measures may serve to clarify the implication of including leisure in a measure of well-offness. Suppose one decided that a definition of poverty should be a well-offness index below a third of the national average of that index. Then the income level at which a family was counted as poor would depend not only on the family size, but also on how much leisure the head and wife had, and therefore on how many hours they were working (or ill or unemployed). It would be assumed that hard working people required more income in order to be equally well off. We are not, of course, proposing such a measure as a standard of eligibility for income supplements, but such a use would present interesting enforcement and incentive problems which we will not treat here.[13]

The basic model of the study is a recursive one with three major types of variables: economic behaviors, economic status, and attitudes. The set of economic behaviors are determinants of changes in family economic status, changed status leads to changes in attitudes (i.e., aspirations and goals and subjective probabilities about the likely payoffs to alternative economic behaviors), and those changes are in turn expected to change the economic behaviors.

The relevant economic behaviors include investing in the human capital of the family, moving where there are better jobs, working hard and "responsibly," keeping in touch with people or organizations that can provide information or help, and using the family resources economically (as distinct from getting control over more of them), and avoiding risks.

Figure 3 gives an idealized version of the model, without specifying whether the variables are levels or first differences, and with some bold assumptions about the direction of causation. Certainly some behaviors are made easier by affluence, and a man's behavior may change his attitude in a process known to psychologists as dissonance reduction (rationalization), but the main forces, we argue, go the way the arrows go.

At this time we have data only from the first observation on our panel, but as subsequent annual interviews are processed we will be looking at the temporal process of the model. For the moment, however, we can look at the static correlations among the three groups of variables, keeping in mind that they represent relations at one point in a process. If the three major groups of variables, economic behaviors, economic status and attitudes move together, we cannot tell whether correlations among them at a point in time should be positive. If there are clear-cut lags in the relations between at-

---

[13] Clearly we have not exhausted the dimensions even of economic well-being. For an extensive discussion of more dimensions of level of living, see S. M. Miller, Martin Rein, Pamela Roby, and Bertram M. Gross, "Poverty, Inequality, and Conflict" in *Annals of the Amer. Acad. of Polit. and Soc. Sci.*, Sept., 1967, pp. 16–52.

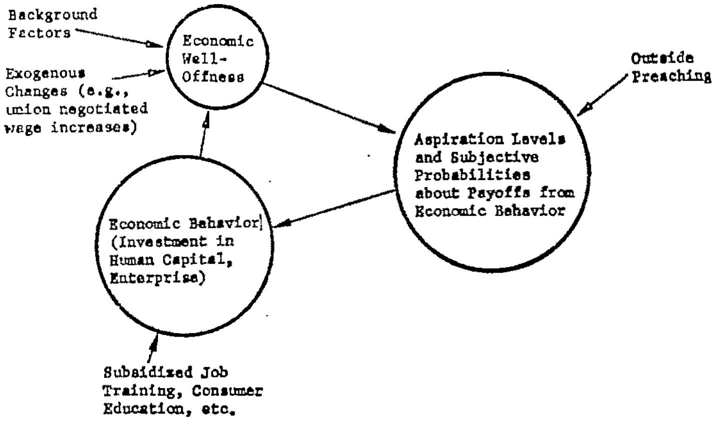

Background Factors
Exogenous Changes (e.g., union negotiated wage increases)
Economic Well-Offness
Outside Preaching
Aspiration Levels and Subjective Probabilities about Payoffs from Economic Behavior
Economic Behavior (Investment in Human Capital, Enterprise)
Subsidized Job Training, Consumer Education, etc.

FIGURE 3
THE THEORETICAL MODEL

titudes or economic behaviors and economic status, then static correlations between either of the former and economic status could be negative and the dynamic correlations between them positive. It is also worth keeping in mind that for some economic behaviors to take place there may be some required level of perceived relative deprivation. We have in mind certain kinds of economizing and moonlighting which may be undertaken when one sees himself as poor but forgone when he begins to see himself as middle class.

We shall restrict ourselves in Figure 4 to three measures of well-offness: (1) money income, (2) net real income/needs, and (3) (net real income/needs) × (leisure); four attitudinal indices cavalierly labeled: ambition-aspiration, sense of personal efficiency, trust in others, and self-reported time horizon and planning; and six indices of economic behaviors: actions to increase money income, connectedness to information and help, behavioral evidence of planning, risk avoidance, actions to increase nonmoney income, and economizing.[18]

The static correlations among these variables are shown in Figure 4. The three small triangular areas along the hypotenuse of the large triangle show the intercorrelations among measures in the same classes, and the three rectangles show the correlations between measure belonging to different classes.

Examining the figure systematically, let us first look at the patterns of intercorrelations among measures of the same class.

The upper left triangle shows positive correlations among the first four income increasing activities, staying connected to sources of information and help, planning ahead, and avoiding risk, but low or negative correlations of those four with the last two: economizing and real earning acts. The last two, in turn seem to go together ($r = .19$). Several interpretations are possible. Because of the way we measured them, some of the income increasing acts and planning acts (time horizon) are things relevant only for those currently employed, whereas do-it-yourself (real earning acts) and economizing are relevant for all families. They also may be the only adaptive activities open to housewives, the retired, and the seriously disadvantaged.

Measures of economic status, we saw in Figure 2, are highly intercorrelated. How-

---

[18] An appendix providing a detailed description of these variables is available from the authors.

	Behaviors						Measures Economic Status			Attitudes			
	1	2	3	4	5	6	7	8	9	10	11	12	13
							(See stubs at left)						
BEHAVIORS													
1. Income Increasing Acts													
2. Connectedness to Information and Help	.08												
3. Evidence of Time Horizon	.21	.18											
4. Risk Avoidance	.12	.18	.27										
5. Real Earning Acts	.02	-.02	.06	-.09									
6. Economizing	-.07	.03	.03	-.05	.19								
MEASURES OF ECONOMIC STATUS	Behaviors → Status												
7. Family Money Income	.35	.17	.25	.36	-.08	-.17							
8. Net Real Income/Needs	.24	.12	.15	.43	-.11	-.24	.84						
9. "Well-Offness" (net real income/needs) x Leisure	.06	.10	.09	.42	-.15	-.23	.72	.94					
ATTITUDES	Attitudes → Behavior						Status → Attitudes						
10. Ambition-Aspiration (Verbal)	.44	.11	.39	.00	.16	.05	.17	.00	-.11				
11. Sense of Personal Efficacy	.19	.15	.15	.33	-.03	.07	.28	.30	.29	.05			
12. Trust in Others	.45	.18	.19	.17	-.01	-.09	.29	.22	.14	.29	.28		
13. Self-Reported Time Horizon and Planning	.17	.09	.18	.20	.05	.03	.19	.17	.14	.20	.59	.12	

FIGURE 4

INTERCORRELATIONS WITHIN AND BETWEEN MEASURES OF ECONOMIC STATUS, BEHAVIORS AND ATTITUDES

ever, when we look at the simplest and the most complex measures—money income versus well-offness—we find that money income explains only one-half the variance of well-offness.

The four attitudinal measures in the triangle at the lower right correlate at levels varying from .07 to .60, the latter spuriously high between sense of personal efficacy and self-reported tendency to plan ahead, which contain common elements, and the former (low correlation) between that same sense of personal efficacy and stated ambition.

The correlations between economic behavior and current economic status (left central rectangle) are substantial for risk avoidance, small but positive for most of the others, but negative with economizing and real earning activities (do-it-yourself). This does not mean that economizing and do-it-yourself will not help improve economic status dynamically, but only that economic pressures to do them are stronger under adversity. Later data on change in status will show whether at any given initial level those who do those things are more likely to end up better off.

In the lower left corner is a rectangle of correlations between attitudes and behavior, the influence assumed to go from the attitudes to the behaviors. The few substantial correlations should not be taken seriously, since they mostly result from either overlapping components in the indexes or overlapping eligibility; that is, items for which only those in the labor force could qualify. The largest real (non-spurious) correlation is between self-confidence (sense of personal efficacy) and risk avoidance, though even here the causation could go in the other direction. Again the last two—real earning activities and economizing—are negatively correlated with some attitudes and not highly related to the others, except perhaps for correlation of .16 between do-it-yourself activity and verbal expressions of ambition. But in general the results are encouraging.

The final rectangle of 12 correlations in the lower right rectangle shows that favorable economic status is indeed associated with favorable attitudes. We are using the term favorable here not in a normative sense but to describe attitudes or behaviors that would seem to lead to higher economic status, as we measure it. Since one of our status measures includes leisure, we cannot be charged with favoring things over time to enjoy them. However, it is part of the assumptions of the model that people who plan ahead, look for better jobs, keep in touch with people who can inform and help them, avoid risk are likely to improve their economic status as we have measured it.

As noted, some of the indexes contain components relevant only for the employed, and since the very young, the very old, and those not in the labor force have much less control over their own status anyway, we reran the matrix of intercorrelations for 35–54 year old people in the labor force. The main results are summarized here without reproducing the whole set of correlations.

The correlations between economic status and the attitude indexes of trust in others and ambition were much lower. Self-reported planning (time horizon) has a higher correlation with income (as it should), but a lower correlation with the attitude indexes on trust in others and ambition. Interestingly enough, the index of income increasing activities was less correlated with income (even negatively correlated with well-offness), and much less correlated with all the attitudinal indexes. Risk avoidance was more highly correlated with income and with attitudes. Finally, the correlation between well-offness and being connected to sources of information and help was lower (perhaps because such connections are more important to the young, the old, and those who are out of the labor force).

The model also allows for exogenous events to change any of the three classes of variables and for background factors to affect them in a more static way. The static data are easily used to examine once more the importance of background factors in affecting the level of present economic status. An index of "background problems" made up of such things as having little formal education, growing up on a farm, correlates in the twenties with each of the three measures of economic status, and in the expected direction with each of the four attitudes (r's from .11 to .18).

One way to take a better look at the relations among the three types of measures—status, behavior, and attitudes—is to measure status not absolutely but net of the effects of background factors. People who engage in striving and adaptive activities should be better off than others with their same background, and it is those same people who should have more confident and ambitious attitudes. As a first step in this analysis, we need a flexible way of eliminating differences due solely to some background factors. Since these factors are unlikely to be additive (any one or two handicaps may be enough), we have used a nonlinear nonadditive model to produce Figure 5, from the cross-section sample, using our most complex measure of economic status.

The criteria used in developing Figure 5 is reduction in the unexplained variance, hence some things such as disabilities which may have an extreme effect on very small numbers of people never come into play. And the introduction of more poor people when the full sample is assembled may well bring some of the other variables into play. As it stands, however, the

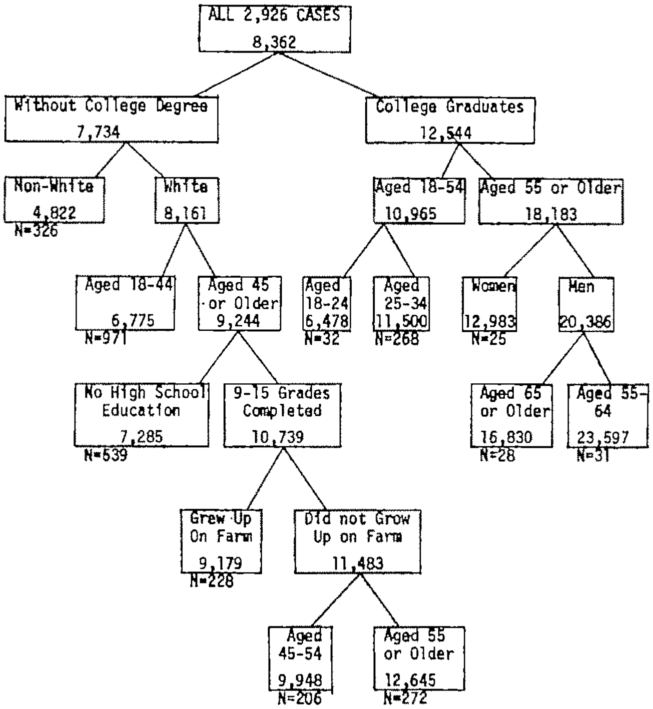

ALL 2,926 CASES
8,362
Without College Degree
7,734
College Graduates
12,544
Non-White
4,822
N=326
White
8,161
Aged 18-54
10,965
Aged 55 or Older
18,183
Aged 18-44
6,775
N=971
Aged 45 or Older
9,244
Aged 18-24
6,478
N=32
Aged 25-34
11,500
N=268
Women
12,983
N=25
Men
20,386
No High School Education
7,285
N=539
9-15 Grades Completed
10,739
Aged 65 or Older
16,830
N=28
Aged 55-64
23,597
N=31
Grew Up On Farm
9,179
N=228
Did not Grow Up on Farm
11,483
Aged 45-54
9,948
N=206
Aged 55 or Older
12,645
N=272

FIGURE 5

WELL-OFFNESS INDEX BY BACKGROUND OF FAMILY HEAD

(Lower Figure in Each Box Is Mean Value of Index for Group)

These final groups account for 36 percent of the variance in the index, using only education, age, sex, race, and whether grew up on a farm.

Other variables unable to account for enough variance to show up but introduced into the analysis were:

education of head's father
region
geographic mobility
disability of head
number of questions that had to be repeated
number of replies that had to be repeated
whether head disfigured
age of head at birth of first child

Well-offness is an index: dollars of net real income times hours of leisure divided by dollars needed by a family of that size and shape. Since the average dollar needs divisor is roughly equal to the average hours of leisure multiplier the index is the same order of magnitude as income. (Before some data corrections, cross-section sample only.)

figure fits what we all know and explains at least as much of the variance as a dummy variable regression with the same variables.[14]

Our final question, then, is whether individual differences in well-offness from the average of one's group (the end-group averages of Figure 5) are associated with any of the economic behaviors or attitudes we have measured. An analysis of these pooled residuals (which average to zero overall, of course) shows that they are related to the following, listed in order of the strength of the relationships:

---

[14] For details of the procedure, see J. Sonquist and J. Morgan, *The Detection of Interaction Effects* (Ann Arbor, Mich., Inst. for Soc. Res., 1964).

Variable	*Sign of Relation*	*Percent of Variance Explained*
Risk avoidance	+	3.5
Economizing	−	3.3
Sense of personal efficacy	+	2.1
Employment problems	−	1.6
Current problems and handicaps	−	1.6
Self-reported time horizon	+	1.2
Real earning acts (do-it-yourself)	+	1.0
Planning acts, revealed time horizon	+	0.9
Connectedness to information and help	+	0.8
Money earning acts	+	0.6
Background problems (grew up poor)	−	0.6
Trust in others	+	0.6
Expressed ambition	+	0.5
Past mobility (different state or region)	+	0.2

There was also a substantial positive correlation with the index of housing and neighborhood quality.

Of course these static correlations prove little about the dynamic relations we seek. Indeed, in the case of economizing the direction may even be different—economizing leading to better economic status, which then allows less economizing. It is also possible, of course, that many of the attitudes and even the behaviors are results of economic status, which allows one the flexibility to avoid risk, makes economizing less essential, and gets attributed to one's own efficiency.

Summary

A number of measures of economic well-offness of a family have been described, implemented with data from a national sample composed of a representative cross-section and a sample of poor families. Their intercorrelations are not so high as to leave us comfortable using the less adequate measures. Correlations of economic status with a number of indicators of economic behavior and of relevant attitudes and aspirations are in the expected directions, though we expect the relation of economizing and do-it-yourself activities to change in economic status will prove to be opposite in sign to their relations to current level of economic status. An examination of current status taken as a deviation from the status of one's peers (in age, education, race, sex, and farm background) shows it to be correlated with both economic behaviors and economic attitudes.

GRADUATED WORK INCENTIVES: AN EXPERIMENT IN NEGATIVE TAXATION

By HAROLD W. WATTS
University of Wisconsin

By this time many people are aware—by rumor, magazine and newspaper, if not from personal contact—of an experiment being carried on in New Jersey on the subject of negative income taxation. I would like to confirm these reports and give an account of the motivation behind this study and the contribution it expects to make.

The first benefit payments were made to an initial group of sample families in August, 1968. By spring, 1969, there will be almost a thousand households receiving, or eligible to receive, income-conditioned benefits of a negative income tax variety. There will be in addition several hundred control families. The payments are now scheduled to continue for at least three years, and the total cost of the experiment (including both research and benefit payments) will be around $5 million.

Experiments of any sort—let alone ones that cost $5 million—are usually considered outside the realm of economic research methodology. Therefore, for reasons of novelty if nothing else, there should be some interest in a more specific description of this enterprise. There is, at least in the minds of its promulgators, a carefully considered rationale supporting the need for such an experiment and guiding its design.

In the discussion below I would like to present this rationale, describe the experiment briefly, and conclude with some observations about the role of experimentation in the further development of our discipline.

I must acknowledge at the outset that this is a joint venture involving personnel at the Institute for Research on Poverty, MATHEMATICA, and OEO. In serving as spokesman for the experiment I cannot take extraordinary credit for any merit or demerit you may see. I am, of course, fully responsible for providing an accurate and balanced description.

Introduction

Economics, as one among the social sciences, has a tradition of concern with the improvement of public policy. Its former name—political economy—can be interpreted as a direct reflection of this concern. But the very structuring of economic relationships and models reflects more substantively than the other disciplines the concern with finding answers to questions that begin, "What would happen if we changed our policy to . . .?" We are continually developing, or adapting, our theoretical frameworks to make them relevant to current problems. And we are also seeking empirical evidence to test these frameworks and/or provide some quantitative precision for them.

For the most part, economists have had to be content with data produced as a by-product of private and public administrative and regulatory activities. In recent years such sources have been augmented by survey data of various kinds. But in both the former and the latter cases the data collected have their range determined by the "accidental" experiments of reality. Sometimes this process provides evidence adequate to answer the question at hand. Sometimes, indeed, it is overabundant. But there seems to be no basis for assuming that sufficient unto the task is the data therefor. There are indeed cases where historical experience simply does not provide direct enough evidence. In these cases—and perhaps only in such cases—the possibility of experimentation ought to be considered.

The experiment in New Jersey—to come to our concrete example—finds its primary rationale in filling a gap in our knowledge of the labor supply function. The revival of interest in direct income redistribution, sparked by the War on Poverty, raises acutely important questions about the effect of redistributive transfers made to households whose current unsatisfactory level of income is mostly obtained through their own sweat.

Static theory does provide some qualitative answers to our questions—specifically that the beneficiaries of a redistribution will tend to reduce their labor supply. But even within that framework the "natural evidence" is inadequate for making a responsible quantitative forecast of the impact of a negative income tax—or family allowance or guaranteed annual income. There are also questions about the basic adequacy of the static income leisure choice model that cannot be satisfactorily tested with available data. Estimation of the labor supply response requires information on the behavior of households with different earning capacities enjoying the same level of unearned income and also households with varying amounts of unearned income facing similar earning opportunities. The usual cross-section turns up an insignificant amount of unearned income in the lower ranges of earning capacity, and what little there is tends to be either extremely transitory (gifts, prizes, inheritances), or property income that is the result of choices highly endogenous to the household's intertemporal allocation of its resources. For both of these reasons it would seem hazardous to project the apparent effects of such income to the case of a wholly exogenous and ostensibly permanent shift in the household's opportunity locus. Similarly, it is hard to find plausible analogues for the drastic change in the marginal trade-off rate between leisure and income that would be exogenously induced by the negative income tax schemes currently under discussion.

We have here, then, I would argue, a case where important public policy decisions may soon be made. Reliable information about both the cost and income augmenting effect of specific redistribution plans is a *sine qua non* of any well-informed choice; and both these depend crucially on the labor supply response to particular features of these schemes that are not well approximated by any phenomena occurring "naturally" in our economic and social system. Given this basic situation, all our cleverness and wit will not draw what we need out of our usual sources of empirical raw materials—which leaves us with a very strong case for inducing experimentally the sorts of changes that will permit us to make the crucial inferences.

When one considers seriously the prospect of singling out a sample of households and giving them cash transfer payments of various kinds and amounts, one immediately has apprehensions about the validity of their actions as an indicator of the response we should expect from the full-scale application of such measures. Misleading artifacts of the experimental situation, often referred to as Hawthorne effects, could (and some are saying certainly will) make the evidence valueless. There seem to be examples, such as the Western Electric experience, that are constantly brought up as reasons for despairing of experimental activities in largely uncontrolled environments. Such doomsayers notwithstanding, social experimentation of this kind is so rare that we really don't know much about these hazards.

Nobody even knows whether three years is a long or short time horizon for the people we are interested in.

The experiment described below has been designed with the threat of Hawthorne effects vividly in mind. Several features of the design are explicitly aimed at minimizing these; and if we knew further ways to reduce them, we would make added adjustments. But there has also been a stubborn (I suppose) refusal, given the depth of ignorance about labor supply, to admit defeat before the fact.

Description

Stated broadly, economists and other social scientists ought to be able to make scientifically-based statements about the social and economic effects of a radical change such as the introduction of a major redistributive program. On the economic side, we can say that such a program will change the "opportunity locus" or "budget lines" for individual choices between leisure and income, spending and saving, investing and consuming. Clearly, the outcomes of these choices are crucial to the performance of our largely decentralized economic system, as well as to the future development (or fortunes) of the choice-makers.

The experiment has been designed primarily to investigate the leisure/work choice, largely because of the relative economic importance we place on it, but partly also because fears about work-effort implications loom so large in the mind of the public at large and the politician in particular. Other economic effects, however, will also be investigated along with a host of noneconomic ones—such as social and political participation, alienation, time preference, family stability, and so forth.

Given our primary concern with labor supply response, the experiment has been limited to households in urban industrial areas which include at least one working-age (18–58) male who is neither a full-time student nor permanently disabled and at least one other family member. The eligible population is further limited to those whose "normal"[1] income places them in the poor or near-poor categories (less than 1.5 times the official poverty thresholds). Experience so far suggests that as much as 80 percent of the eligible households will have normal incomes above poverty; i.e., between the poverty threshold and 150 percent of it. In Paterson-Passaic, 41 percent of the eligible families are Negro and a disconcerting 44 percent Puerto Rican, leaving only 13 percent of available households to represent the non-Puerto Rican whites—a group which, according to most recent national survey data, still comprises a majority of the poor even in urban areas. This is a matter of some concern since we would like, at the very least, to have a sufficient portion of the sample in the "native white" category to detect any ethnic differences in response there may be. The composition of eligibles in Paterson-Passaic is given in more detail in Table 1.

The group we have chosen as our primary target for experimentation—intact families with at least one working-age male—bulks large in the total poverty population and is, also, in the main, outside the eligible categories of existing wel-

---

[1] As used above, "normal income" refers to an empirical approximation to a long-run income concept such as Friedman's Permanent Income. We have been developing regressions for describing the average "relation" between family income and a fairly eclectic set of household characteristics. This relation has been fitted so as to yield a good approximation at the low end of the income distribution. We propose to use an interpolation between (1) a household's predicted income by this regression and (2) its actual income over the most recent year as reported on the screening interview, as our normal income. Turned around the other way, we are attempting to adjust the short-period income figure, that we measure directly, according to its consistency with other, non-income measures of the households' income potential.

TABLE 1

PERCENTAGE DISTRIBUTION OF ELIGIBLE HOUSEHOLDS IN PATERSON-PASSAIC

	Poor	Near Poor	Total
Family size			
Small (2–3).....	3	17	20
Medium (4–5)...	5	33	38
Large (6+)......	10	32	42
Race			
Negro...........	6	35	41
Puerto Rican....	9	35	44
Other white......	3	10	13
Not ascertained..	0	2	2
Total............	18	82	100

fare programs. It is perhaps fair to say that this "working poor" group is excluded from such programs precisely because of fears about their labor supply. From the experimental point of view, these are the households most likely to have observable changes in their labor supply, because these are the household heads with no income alternative up to now except full-time work if they can get it. Female and aged heads have historically contributed less to the labor market.

Homogeneity is another criterion that has influenced our selection of the experimental group. Clearly one can carry homogenization to ridiculous extremes and end up with a group so narrow that the outcome is of no interest. But on the other side, an experiment of this sort is such a novel undertaking that it needs all the help it can get. There are no obviously outstanding economies in trying to cover the whole population at a single shot, so we have taken the prudent course of applying this largely untried methodology to a limited group. We have achieved important simplifications by excluding certain groups. In some cases the excluded groups —the rural poor is a primary example—should be candidates for further experimentation (indeed plans are currently under way for such an experiment). In this sense the New Jersey experiment should be regarded as dealing with an important, and hopefully tractable, part of the problem, but requiring supplementation from further experiments. The main groups we have excluded are unattached males, female heads, aged heads, and all rural poor. We are also giving only perfunctory treatment to the intractable (but important) self-employed.

Households meeting our criteria and selected for our sample are either assigned to a "treatment"—which is one of a range of eight linear transfer schemes—or to a control group—which, of course, is merely observed, being given a nominal payment for the time taken up by each interview. Any specific linear scheme can be described by two numbers—the tax rate r, and a scalar, g, which indicates the level of the maximum benefit or income guarantee. The dollar amount of the income guarantee for a specific family depends on household size and on the value of g for the plan assigned to that family. Table 2 shows the basic guarantee schedule for families of size 2 to 8+ (single individuals are not, of course, included in the sample). The actual guarantee, then, is the product $g \cdot G(n)$. The household's net benefit is calculated as the difference between the maximum benefit $g \cdot G(n)$ and the product of the negative tax rate and income. The net benefit diminishes and

TABLE 2

BASIC GUARANTEE SCHEDULE

Household Size, n	Basic Guarantee, $G(n)$
2	$2000
3	2750
4	3300
5	3700
6	4050
7	4350
8 or more	4600

finally vanishes at the break-even point $g/r \cdot G(n)$.

Current plans call for eight different combinations of g and r, specified in Table 3. In these plans the tax rate ranges from .3 to .7, and the guarantee level from 50 to 125 percent of the basic schedule in Table 2. Table 3 also shows the break-even levels as fractions of the basic schedule. The basic guarantee schedule is scaled for four- and five-person families at the level of the official poverty-threshold levels, but it departs somewhat for larger and smaller families. Only the first seven plans are currently in operation in Trenton. We plan to add the last plan in our second and third cities.

As with any tax system, definitions of the unit and of the income of that unit are important components of the program. Being important, it could well be argued that some experimentation should be done to ascertain more precisely the effects of alternative definitions. Again, we have chosen not to complicate things further in this way, but to use our best judgment in specifying a single set of rules, given the precedent and experience of both the positive tax system and existing welfare programs.

In defining the eligible unit, the decision was made to include (besides the spouse) any child or stepchild or a descendant of any child or stepchild if such a person is either living with the head or derives more than half his support from the head. In addition, any other person who both lives with and derives more than half his support from the head is included. This definition conforms fairly closely to the economic notion of a family decision-unit and is also close to the census-type family unit underlying the statistical tabulations we are all familiar with. It does differ from the positive income tax definition in that our basic unit is the family, while in the positive income tax it is considered the individual (this has only been partly qualified by joint filing—which is optional).

TABLE 3

EXPERIMENTAL COMBINATIONS OF TAX RATE AND GUARANTEE LEVEL

	r	g	g/r
I	.3	.50	1.67
II	.5	.50	1.00
III	.3	.75	2.50
IV	.5	.75	1.50
V	.7	.75	1.07
VI	.5	1.00	2.00
VII	.7	1.00	1.43
VIII	.5	1.25	2.50

Because of the peculiarities of an experiment, as compared with a full implementation of such a law, some additional features must be included to eliminate the possibility of uncontrolled additions to the experimental sample. Consequently, no new individuals who join the eligible household after initial eligibility has been established will be included for our purposes (except children born to an eligible female and any minor child who joins the unit after an initial waiting period of six months). Initially eligible unit members who join the armed services or leave the United States or are institutionalized (except those voluntarily institutionalized and who remain dependent upon the tax unit for support) will cease to be unit members while in such status. Arguments in favor of making persons marrying an eligible person also eligible can be made, but we have decided against this, because such a provision would produce (in our experimental setting) an effective dowry which could distort normal mating patterns purely as a result of the artificial setting.

It is impossible to simulate the family composition incentives of a full-scale program in the experiment. The best we can do is to be as neutral as we know how relative to the *status quo*. We may be able to get some general information about how sensitive such choices are, but nothing more.

Before moving on to the income definition, it should be explicitly mentioned that the definition of the unit bears on the level of benefit calculation in two ways: (1) through its implications for family unit size and (2) through its implication for the level of income—because income received by all members of the unit must be included in the income calculation.

Items added to the definition of current taxable income for the experiment represent a substantial modification of the definition of income used in the positive tax system. Our concept will include: (*a*) any pension or annuity (including social security); (*b*) all prizes and awards; (*c*) all life insurance proceeds over $1,000; (*d*) gifts, support payments, and inheritances over $100 from persons outside the family unit; (*e*) all interest on governmental obligations; (*f*) damages, insurance payments or workman's compensation due to injury or sickness including wage or income continuation; (*g*) all dividends; (*h*) fellowships or scholarships, including value of room and board supplied without charge, to the extent that such stipend exceeds the costs of tuition, fees and books; (*i*) income from trusts and estates; (*j*) gross rental value of owner-occupied housing or other quarters occupied rent free; (*k*) capital gains counted in full as income, and losses to be deducted to the extent of gains received during the period of the experiment; (*l*) payments received from (*i*) unemployment compensation, (*ii*) strike or supplementary unemployment benefits, (*iii*) Social Security Survivors benefits, (*iv*) veterans disability benefits, (*v*) training stipends.

Specific exclusions are the following benefits and others for which the size of the benefit is based on demonstrated need: (*a*) aid to the permanently and totally disabled; (*b*) old age assistance; (*c*) aid to the blind; (*d*) aid to families with dependent children; (*e*) general welfare and public assistance.

An additional word should be said about (*d*) and (*e*). When our experiment was planned, New Jersey had no Unemployed Parent provision in their welfare programs. We, therefore, did not have to worry about welfare payments confusing our observations, since the only significant welfare program was AFDC and we were not concerned with households headed by females. Time caught up with us, however, and starting January 1, 1969, New Jersey is instituting an AFDC-UP program—the benefit level of which is higher than all but our most generous plan. We have been forced, therefore, to reconsider our previous cavalier decision simply to ignore competing welfare payments. We have adopted the following rule:

> Any family in our sample who becomes eligible for AFDC-UP benefits will remain in our sample. During any month for which they elect to receive those benefits, they cannot receive any payment from us. They do remain eligible to return to our plan whenever (and as soon as) welfare payments stop. [As New Jersey's welfare laws are written, this provision does not affect a household's final disposable income—New Jersey would offset any amount we paid to welfare clients anyhow. It does, however, ensure that we are not simply reducing New Jersey's welfare outlay in return for observations of no use to us.]

Deductions are limited to the following: (*a*) businessmen and independent contractors may deduct costs of earning income; (*b*) mortgage interest and real property taxes may be deducted in full in computing income from owner-occupied housing; (*c*) payments of alimony to persons outside the tax unit may be deducted; (*d*) finally, full credit (reimbursement) will be allowed for all federal income taxes paid or withheld, and reductions in benefits will be made to offset any tax refunds received. Thus the recipient's unit will be insulated from the regular tax system until its income reaches the level at which the unit will fare better under the positive tax system.

This definition aims at including most components of "total family income" as measured for census purposes. This broadening seems to be called for in a program aimed at relieving genuine need as compared to tax-exempt status. Some of these provisions will have little if any application in the experiment, but they do need to be spelled out to take care of the surprising exception.

It may be useful here to outline very briefly the sequence of events involved in a family's becoming enrolled and operating under one of the experimental plans. The first contact will be through a screening survey to identify a group of eligible households. Very short interviews are administered to a random sample of households in the poverty areas aimed at eliciting just enough information to determine the household's eligibility for inclusion in the experiment. Our experience in the poverty areas has been that 10 percent or less of the screened households are eligible. A family that is eligible will be contacted again (if it is selected)—in a second-stage stratified sampling process—for enrollment in one of the experimental plans or the control group. It will first be interviewed at some length (the pre-enrollment interview) to provide base-line data (pretreatment measurements), and shortly thereafter will be visited by enrollers who will explain the particular plan assigned to the family and attempt to secure their agreement to participate. If they do agree, they are given their first benefit check immediately and another one every second week. Every fourth week they are required to report their income for the preceding four-week period. At the time of enrollment they are also given a schedule which shows in some detail the size of the benefit they are entitled to at each level of reported income.

So long as the household continues to report monthly and send in the relevant pay-stubs, they continue to receive benefits for a three-year period—which come through the mail in the form of biweekly checks, drawn upon the account of the Council for Grants to Families.[2] Their continued eligibility does not require that they stay in New Jersey. We are prepared to continue benefit payments wherever they may go, within the continental U.S., provided only that they continue to send in their income reports.

If their income rises above the break-even point, they are of course entitled to no net benefits, but they will remain eligible should their income subsequently decline. In an effort to maintain contact with households receiving no benefits we do continue to give a minimum payment for their trouble every time they send in an income statement, and allow them to change to a quarterly income-reporting period until such time as they are again eligible for benefits. Since it is undesirable, and probably not even feasible, to get a contractual agreement that limits their future choices, we are mostly dependent upon making it financially attractive for them to continue to participate. The related problem of competing welfare payments has been dealt with above.

Each household in the experiment, including control households, is interviewed every three months. These interviews last about an hour and will, except for the same short section on labor force activity every time, vary from quarter to quarter over the thirty-six month period, depending on how often data have to be gathered on specific questions. (Children's performance in school, for instance, will hardly need to be checked every three months; and past biographical data need to be collected only once.) These interviews provide the main source of data for analyzing both economic and noneconomic responses to the treatments. Along with the pre-enrollment interview and the monthly income and family-size reports, therefore, there will be a large quantity of longitudinal evidence on each family.

---

[2] A wholly-owned subsidiary of MATHEMATICA.

Cooperation with the quarterly interview is not a requirement for receiving benefits; and, of course, the control group does not receive benefits anyway. Consequently there are the usual problems of maintaining a panel of respondents over a three-year period. Cash payments will be made for each interview and additional efforts made to enable us to maintain contact with those who change residence.

There is an inevitable conflict between our desire to gather substantial amounts of information and our intent to place no welfare-type restrictions on the use of benefit payments or, indeed, on the activities of the recipient households in general. The only requirement for benefits is prompt and accurate reporting of income and family composition. We are, therefore, making every effort to keep the administration of payments separate from the collection of data both in fact and in appearance. (There are indeed two nominally separate organizations dealing with the households.) No doubt either the actual behavior or the response to the interview will still be distorted to some extent by this process, but we expect that such bias will become small over time as the sample becomes accustomed to the routine.

As currently operated, there is a local office in each experimental site. This office serves during the screening process as a headquarters for a large interviewing crew. It also provides a base for efforts to establish contacts and, with luck, a degree of rapport with local figures of authority. Such contacts can provide useful assurance to experimental households at the time of enrollment. After the experiment is under way these offices serve to answer questions, to receive complaints of nonreceipt of benefits, etc., to assist the experimental households with their income reports, and to investigate failures to report.

Supervision of operations is centered in Princeton, under the responsibility of MATHEMATICA, Inc. Housed there are the claim and disbursement activities, and the data coding, editing, and processing functions, along with the general management operations.

The operations at the Institute for Research on Poverty are limited to planning, design, questionnaire formulation, and analysis (all of which are carried on jointly with people at MATHEMATICA), plus the overall financial management and ultimate responsibility.

As mentioned above, households will be drawn from a stratified sample of eligible households and assigned to specific experimental variants or to the control group. Defining and solving the problem of assigning households in an optimal manner is a significant study in its own right. We have stated the problem as one of maximizing the informational output per dollar of cost. The "information" being maximized is specified as increasing with the inverse of the expected forecast error for a projection of total disbursements under a full-blown implementation of a similar transfer program. Taking the static or no-response estimates of such disbursements as given with no error (or at least irreducible error), we are mainly concerned with the incremental costs produced by induced changes in earnings and other income sources.

It is possible, by using suitable restrictions, to specify the marginal information value of an additional household with a given "normal" income assigned to each of the experimental groups. The cost of such an assignment depends on the household's normal income, its size, the generosity of the particular transfer plan, and, alas, the very same labor force response the experiment is intended to illuminate. We have been able to use a range of guesses about response in order to carry out the optimization. Our best judgment is that most of the uncertainty lies in the level of these responses rather than in their relative magnitudes. If this turns out to be correct, our optimization model will provide a good pattern of assignment. We can hedge our limited budget by making a conservative guess about the costs, and then extend the sample (and/or the time period) if actual responses are enough like our a priori estimates.

With the several parametrically-defined variants of the treatment, we shall be able to utilize regression techniques for exploring the response to tax rates and guarantee levels. In an important sense the estimates will be based as much upon differential response among the experimental variants as between each of these and the "unpaid" control group. This provides an effective means of minimizing biases from Hawthorne effects—if such effects are additive, they do not affect intergroup comparisons among the experimental groups.

By taking advantage of the longitudinal nature of our data, we can also observe the time pattern of any adjustment induced by the experimental treatments. This may give us some further clues about the validity of the observed response as well as simply extend our knowledge of such processes. Our first approximation holds that the most reliable indication of the steady-state consequences of redistribution will be produced in the middle of the experimental period—after the novelty has worn off and the adjustments completed and before the anticipation of the end of the benefit period contaminates the results.

With the experiment described above in mind, it is worth considering the distinction between an experiment and a "demonstration." In the former there are well-specified hypothesis tests or estimation tasks to be carried out from the experimental evidence. Building from available knowledge, the object is to extend that knowledge in quite specific ways. In the latter, the important questions have either been answered, or not asked, and the object is to convince a wider public that a specific preselected program won't sour the milk. The emphasis is on extending the recognition or acceptance of a policy or program and not on extending the knowledge needed for selecting alternative policies or programs.

One cannot readily carry this useful distinction to the point of identifying pure types—many "demonstration" projects can and do extend our knowledge. Similarly, an experiment such as the one described above is producing experimental variation in some dimensions and not in others—and in this latter category may be said to be providing a demonstration of feasibility. Many of the administrative details such as monthly income reports, nonstandard income definitions, etc., will be "demonstrated" as workable or abandoned. But despite the fact that experiments and demonstrations usually occur in some sort of mixture, they should not be treated as synonymous. Experimentation has a great potential as a source of scientific evidence on social and economic behavior. As such it merits consideration as an addition to our research toolbox.

The potential of "demonstrations" seems much more doubtful.

Conclusion

At the present moment there is little to report of a substantive nature beyond the fact that almost all of those who have been invited to participate in the payments program have chosen to do so, and so far most of them are reporting their incomes promptly—and moreover still have an income to report. No immediately apparent outbreak of hives or plague of locusts has been induced. From a scientific point of view the most encouraging result is that with enough patience, understanding, and luck, it is feasible to initiate a social experiment that substantially (1) alters the alternative choices of some households and (2) introduces several graduated treatments as well as the null or "control" treatment. At the outset, such an undertaking within troubled center-city areas appeared formidable—and still does—but at present we remain convinced that meaningful experimentation of this kind can be carried out.

Without generalizing from a sample of one, I am encouraged to report that this project—whether because of its importance, novelty, or challenge—has succeeded beyond anything in my experience in bringing a variety of academic scholars from diverse disciplines into concerted, prolonged and fruitful contact. Interdisciplinary activity has a rather bad name and may richly deserve it. But there are conditions under which it can be successful. From the point of view of the social scientist this example also provides a most welcome assurance that it is possible in yet another way to make a substantial contribution toward solving urgent social problems from our so-called "ivory tower."

DISCUSSION

CHRISTOPHER GREEN: The three papers presented this morning report on projects which have just begun—or, at least, are far from completion. The results reported—if any—are preliminary or incomplete. Instead of discussing the broad questions of methods and objectives in these studies, I shall consider a few isolated points which I hope are relevant to some aspects of the topics of the papers.

First a word about the Budd-Radner paper. The construction of a comprehensive income size distribution by synthesizing Current Population Survey (CPS) data and information from income tax returns should fill a serious void created by inconsistent estimates of the size distribution of income based on existing surveys of family income. For example, when I was estimating poverty income gaps in an attempt to calculate the cost of some negative income tax plans, I found that the $15.0 billion poverty gap I calculated using CPS data for 1961 was double the $7.5 billion gap I calculated using consumer unit income data from the Bureau of Labor Statistics' Consumer Expenditure Survey for 1961. Of course, the major purposes of these two surveys are quite different and as a result their definitions of the family unit differ, which undoubtedly accounts for part of the difference in the poverty gap estimates. I am reasonably certain, however, that relatively greater underreporting and nonreporting of income in the CPS was an important factor producing the widely differing estimates of the poverty gap. The poverty population adjustments for 1964 made by Budd and Radner provide additional evidence that CPS data also overstate the number of persons and families living below the Social Security Administration poverty lines.

The aspect of the Morgan-Smith paper which especially aroused my interest is their construction of a well-offness index as a measure of economic status. The well-offness index is composed of two subindexes: one, the ratio of real income to a needs standard, is an index of the "adequacy of the family's command over resources"; the other is an index of leisure consumed by the head of the household and his spouse. In order to combine these two partial indices of welfare into a single index of well-offness, Morgan and Smith assume that "leisure and income are substitutable commodities in a preference map of rectangular hyperbolas."

Reference to a preference map suggests that the measure of well-offness is also a measure of ordinal utility. We may ask, then, what combination of hours of work or hours of leisure maximize the household's well-offness (utility), given household size, the wage rate(s), and the amount of nonemployment income. In the simple case where only the head has the option to work and the family has no nonemployment income, well-offness is maximized when the head works exactly half of the hours for which work and leisure are substitutes no matter what the wage rate. Happily, Morgan-Smith assume there are sixteen hours a day for which work and leisure are substitutable, so well-offness is maximized by an eight-hour day. Even when some other utility functions are used, for example, a Cobb-Douglas function, the wage rate does not appear to matter. However, when nonemployment income enters the picture the wage rate does matter with well-offness maximized when $H = T/2 - P/2w$; where $H =$ annual hours worked, $T =$ total hours for which work and leisure are substitutable, $P =$ nonemployment income, and w is the wage rate.

The relative noninfluence of the wage rate on annual hours worked under the maximizing assumptions is consistent with empirical studies of the individual labor supply curve which indicate that labor supply of primary workers is highly wage inelastic or even slightly negatively sloped. However, the Morgan-Smith well-offness ratio does imply that when a family has a considerable amount of nonemployment income and a very low wage rate or if it had transfer income from some form of negative income tax, the reduction in work effort by utility-maximizing workers may be quite substantial.

When both spouses have the option to work it is important how the leisure hours of husband and wife are integrated into the well-offness index. Maximization of well-offness leads to meaningful economic results when leisure hours of the head and spouse are multiplied within the well-offness index, but not when they are added together. The husband-wife model indicates a role for the relative wages of the spouses and also suggests that the utility maximizing household will substantially reduce its hours of work if it is eligible to receive negative income taxes.

Watts's paper reports on the New Jersey negative income tax project. The main objective of the New Jersey investigation is to learn more about the labor supply effects of negative taxes. I should like to make two comments about negative income taxes. First, it seems to me that it is not only important to estimate the labor supply response to a specific negative income tax plan, but also to determine who it is who is responding by reducing hours. Suppose most of the reduction in hours of work is in the form of reduced moonlighting by primary labor force participants or reduced hours of work or reduced labor force participation by secondary workers who are not usually expected to work (e.g., school age persons, mothers with children, or aged persons). While the output effects may be no less important whoever reduces work effort, it seems to me much less upsetting when the labor supply response comes mainly from persons who are working more than some "socially accepted average" number of hours. (The socially accepted average work week for a primary worker might be a forty-hour week, while the socially accepted average work week for a secondary worker ranges upward from zero depending on who the secondary worker is and what are his or her circumstances.

My second comment about negative taxes concerns the means by which a negative income tax device might be implemented. Earl Rolph has proposed converting the present $600 personal exemptions of the positive income tax into tax credits amounting to, say, $400 per person. "Refunds" would be made to tax filers whose tax credits exceed their tax liabilities. The income tax base would be increased by the elimination of the exemptions and expanded further by making some currently excluded income such as transfer income and all of capital gains taxable and making reforms in the area of personal deductions. The whole package would be financed by applying a flat income tax rate in the range from 25 to 30 percent to the broadened tax base.

I find the Rolph proposal appealing. It is the best integration of negative and positive income taxes into a single system of which I am aware. While I attach some weight to the argument that a Rolph-type plan is politically unfeasible because Congress will balk at any general tax reform I find the counterargument that in the next few years any type of negative income tax plan is likely to be politically unacceptable to be just as weighty. Moreover, I would rather wait ten or fifteen years to get a well-constructed negative income tax device than to have us adopt a plan in half that time which is little more than an extension of public assistance. My feeling is that the case for a negative income tax now rests as heavily on the need to achieve greater equity in the overall tax-transfer system as it does on further closing the poverty gap. Three years ago we proponents of negative taxes emphasized the inadequacy of welfare payments and the need for a "new front" in the War on Poverty. Today it is important to recognize that while in many areas of the country public assistance is as inadequate today as it was three or four years ago, in other areas the payments have increased to levels with which no politically feasible negative income tax can compete. For example, in New York City a family of four receives anywhere from $3,600 to $4,000 under the Aid to Dependent Children Program. The improvement in achieving adequacy in New York City and other metropolitan areas has been at the expense of equity—equity not only between the poor themselves but between some of the poor and low middle income families. While Buchanan and others are probably right that our society is not enthusiastic about the goal of redistributing income, I think our society is concerned about horizontal equity; that is, equal treatment of equals. As economists we have emphasized the importance of achieving horizontal equity on the tax side. Now we need to go further and achieve horizontal equity with respect to both the taxes we pay and the transfers we receive.

Harold W. Guthrie: A former colleague of mine remarked at last year's annual A.E.A. meetings that there are just two things about the distribution of income that economists do not understand: poverty and affluence. He was at least partly correct, but we are beginning to understand more as attested by the papers in this session. The papers are progress reports on very ambitious attempts to increase our understanding of these issues. The Budd-Radner paper is concerned with the accuracy of our factual knowledge of income distribution; the Morgan-Smith paper proposes the elements of a new theory of measuring welfare; and the Watts paper outlines an experiment designed to test an important policy proposed to alleviate poverty.

The papers have some common characteristics. Each concerns an interesting aspect of the total problem of income distribution and considers poverty within this broad context. Each promises more significant research results than it is now prepared to deliver. The fact that a session of our meetings is devoted to progress reports of long-term research projects rather than completed projects again may reflect the state of our ignorance of poverty.

The painstaking work being done by Budd and Radner illustrates one of the very difficult problems faced by researchers seriously concerned with data on income distribution. The available data on income are wrong in the same sense that all economic data are wrong. One who pursues the data in all of their complexity of varying types of income units, the lack of agreement between aggregates and estimates from cross-section data, the great variety of sources of income, soon finds that he has learned, perhaps, more than he wanted to know. Budd and Radner are persevering expertly in seeking the sources and magnitudes of error.

There is a danger, however, in the very detailed kind of analysis they are pursuing. Given that all data are approximations we need criteria to tell us when the last nit has been picked. How many decimal points of accuracy in the percentage distribution of income do we need?

There is also a promise in their work of a firm base for more definitive measurement of errors in income distributions. The best procedure so far developed for measuring bias (the difference between a true value and an estimate) as opposed to error (the difference between two estimates) is the validation survey. In the best of validation surveys, the true value of a variable, say wages received, is known along with the observed value for identical analytical units. It is most unlikely that we will ever resolve all differences in estimates until we undertake validation surveys of income by source that will allow measurement of response biases. Budd and Radner's work will be valuable in identifying the types of income most subject to response bias.

The main thrust of the project reported by Morgan and Smith is the measurement of the dynamics of poverty. At present we do not even know how many people in poverty at a given point of time are in that condition permanently or only temporarily. The results of their project so far do not tell us much about poverty but they should be applauded for beginning with the search for a theory to accompany their efforts at measurement.

Whether or not the theory they propose has any special merit with respect to answering questions or developing public policy concerning poverty remains to be seen. In its present stage of development my first impression is that the theory is oriented more toward middle class behavior than toward the condition of people in poverty. We see leisure as a good with intrinsic utility because we work. In the absence of work, leisure has little meaning. It seems to be at least confusing, if not misleading, therefore, to measure with equal degrees of well-offness the evening hours of an employed person and the evening hours of an unemployed person.

The inclusion of nonmoney income in the form suggested is also inconsistent with thinking of poverty as a social ill. Are poor people better off than others because they have poor relatives with whom they can crowd into poor housing and thereby save money? Are poor people better off if they receive food and clothing as charitable contributions or public support? The do-it-yourself component probably serves to widen the gap of well-offness between those in and out of poverty because physical and human capital are required to save substantial amounts of money in this way. This component may well be relevant to people in poverty, but some of the others are at least questionable.

The Watts paper has some unusual values for us as economists. It describes what may well be a "first" for our profession. Large expenditures for experiments in the physical sciences and war games are so common now as to lack any news value. It is good news, indeed, that $5 million is being devoted to an experiment aimed at measuring the operating characteristics of an important public policy.

One final comment on the negative income tax proposals now widely discussed in both public and professional press: the general issues relating to income maintenance programs will continue to be important elements in an evolving public policy. They are important enough to be included in our economics classrooms.

LESTER C. THUROW: The large reduction in poverty caused by the Budd-Radner adjustments raises several disturbing problems. First, it is based on the assumption of equal underreporting for each individual with income of a particular type. If farmers report only 59 percent of their income, then each farmer reports only 59 percent of his farm income. The assumption of equal underreporting is probably not true. In the case of farm income, there are large differences among states in the ratios of USDA estimated farm income to farm income as reported on tax forms. If different states have different amounts of underreporting, then different types of farmers undoubtedly have different amounts of underreporting. Perhaps state differentials could be used to do a more sophisticated allocation of underreporting. Given the states with large amounts of underreporting (Texas and California), I would guess that underreporting for larger farmers is higher than for small farmers.

To understand the significance of the large reduction in poverty, the data should be broken down to quantify the impact of different factors on the prevalence of poverty. How much of the reduction in poverty can be ascribed to underreporting of farm incomes, how much to the underreporting of public assistance, etc.?

I suspect that farm income must be the major element causing changes at the bottom of the income distribution. If only 59 percent of farm income is reported, a farmer who reports an income of $2,000 is blown up to an income of $3,390. According to the CPS, the median income for farmers was $4,841 in 1966 and for nonfarmers it was $7,582. When farm incomes are blown up, farmers earn $8,205 or more than nonfarmers. This certainly reverses a widely held belief that is at the basis of many public policies. The authors should calculate a distribution of farm income based on OBE control totals. From their data farmers should be making much more than parity rather than less than parity. If true, they are brewing a real case of political dynamite.

I would also like to know what their adjustments do to black-white differences. I suspect that they become larger, but perhaps not.

Let us assume that the Budd-Radner income distributions are correct. How should this influence social policies. Should we reduce the number in poverty or change the poverty line? The latter is a relevant question, since the American poverty standard is not a minimum survival line. It is a socially agreed upon norm. When it was set, a definite number of people were believed to be below it. The number below had something to do with setting the norm. Perhaps we should go back to 1963 and set the poverty line at a level which would place the same number of people in poverty in 1963 as were then believed to be in poverty. There are no "technical" answers to these questions, but they are questions that must be answered when "technical" adjustments affect the results as much as these "technical" adjustments do.

I am skeptical about the Morgan-Smith index of well-offness. I do not believe there are any observable signals that can be used to determine when different families feel equally well off. You and I could sit down and reason together or two philosopher kings could sit down and reason together, but none of us would be able to determine whether we were equally well off, better off, or worse off. Well-offness in the utility or physic sense is in principle unobservable.

Individual perceptions of well-offness (cardinal utility?) are irrelevant from the vantage point of social policy. As economists we search for Pareto optimums, but as a society we make interpersonal comparisons and search for the optimum optimorium. We do not need a measure of when families feel equally well off, but a measure of when society thinks families are equally well off.

Leaving this fundamental objection aside, let me turn to their actual efforts to measure well-offness. First, it should be pointed out that the Orshansky definition of poverty is in essence an application of the Friedman approach to measuring equal well-offness. She too assumes that people are equally well off when they consume the same proportion of their income on food expenditures. The only difference is that she uses a simple average rather than a regression to obtain the standard food consumption propensity and then works backwards. She finds a minimum food budget necessary to escape poverty and then multiplies it by three (the inverse of her food consumption propensity) to determine poverty lines for different families. The families are then supposed to be equally well off when they are at the poverty line.

The Friedman approach also assumes a rather peculiar utility function. Each individual has the same utility function. In its reduced form, its only argument is the ratio of income to food expenditures (Y/R) (see equation(3)). The utility of all nonfood commodities must be proportional to the ratio of income to food expenditures. This may look like

$$U = f\left(\frac{Y}{R}, X\right) \tag{1}$$

$$U_X = b_1 f\left(\frac{Y}{R}\right) \tag{2}$$

$$U = f\left(\frac{Y}{R}\right) + b_1 f\left(\frac{Y}{R}\right) \tag{3}$$

Where:
U = utility
Y = income
R = food expenditures
X = nonfood consumption
U_X = utility from nonfood consumption

your utility function, but it does not look like mine. Perhaps it is the social utility function.

Different types of necessities should also yield the same ranking. From the correlations between housing expenditures and food expenditures, I suspect that a very different ranking would be gotten by using housing expenditures.

Finally, I simply do not understand the rationale of the index of well-offness used in the paper. The index is income divided by needs multiplied by hours of leisure. An elasticity of substitution of income for leisure of 1 is as good as any other arbitrary assumption, but I do not see how it leads to the measure in the paper. The paper does not give a derivation and I cannot generate an acceptable one.

There are two comments I would like to make on the negative income tax experiment: (1) it was a major mistake to ignore female headed households; (2) the experiment should have included a group that were given a straight work incentive.

Ignoring female headed households means the experiment will be unable to estimate how much the negative income tax will encourage people on welfare to work. Theoretically the negative income tax should discourage those not on welfare from working and encourage those on welfare to work. Society is interested in the net impact. This experiment may yield negative work effort results (indeed it should); yet the introduction of the plan might have positive work effort results. The experiment is only designed to ask half of an interesting question.

Ignoring female headed households is a mistake for another reason. When AFDC was initially established, it was deliberately designed to discourage female household heads from working. At that time society thought the social benefit of having a mother at home was greater than the social benefit of having a mother in the labor force. As I read the entrails of the public discussion, this value judgment has changed. Society now thinks that the social benefits of a self-supporting mother exceed those of an "at home" mother. Value judgments may have changed because of the rising participation rates for women in the labor force of all social classes, but regardless of the reason they have changed. Thus the incentive effects on female household heads is probably politically more important than the incentive effects on males.

The negative income tax experiment should also have had a control group which was given a transfer payment equal to some fraction of whatever income the individual was able to earn himself. In this way we could have found out the impact of wage rates as well as lump-sum transfers on the labor supply function. The desire for price as well as income effects is not just idle academic curiosity. If the negative income tax does have negative work effects, they could be offset by public work incentives. An optimum welfare program may well be forced to include both a negative income tax and a work incentive plan. Estimates of price elasticities for work effort are as important as estimates of income elasticities.

STRATEGIC THEORY AND ITS APPLICATIONS

THE PRINCIPLE OF "FISCAL EQUIVALENCE": THE DIVISION OF RESPONSIBILITIES AMONG DIFFERENT LEVELS OF GOVERNMENT

By Mancur Olson, Jr.
University of Maryland

Though economic theory provides a good basis for decisions about which functions ought to be performed through free markets and which by collective or governmental action, it does not tell us what type of government or institution should perform those activities that require collective action. It does not tell us whether a function should be performed by a local, state, or central government, by an *ad hoc* authority or organization, by an international organization, or by some other type of institution. Neither does any other discipline have a theory which can be of general use here.

I

Perhaps in part because of the lack of a good general model, there is a wide range of thought (or ideology) on this subject. There are advocates of world government, of unitary national government, of federalism, of "states rights," of metropolitan government, and of maximum possible reliance on the level of government that is (physically) closest to the people. There have also been advocates of "functional" units of governments, which have responsibility for a particular function, occupation, or industry rather than a given geographical area. This point of view was characteristic of many guild socialists, syndicatists, pluralists, and believers in the corporate state, and is still part of the official ideology of the Gaullist movement. The debates on this subject have lately taken on a new twist as the "new left" has forsaken the "old left's" emphasis on central planning and nationalization in favor of "participatory democracy," which appears to involve a degree of decentralization at least as great as the old right or classical liberals have advocated. The modern demands for what might be called "black power separatism" or separate governments in Negro ghettos have added yet another approach, which we shall give special attention here.

Just as opinions on this subject are diverse, so is the practice, even within the United States. There are now, by one count, about 80,000 governments in this country. The New York metropolitan region alone is supposed to have 1,400 governments. The typical citizen is not only under the jurisdiction of national, state, county, and city governments, but sometimes also subject to a metropolitan transport commission, a port authority, a sewage or sanitary district, a Soil Conservation District, a pollution control district, a school district, an airport commission, or a metropolitan planning commission or council of governments. Many of these institutions are functional governments or other authorities or organizations that have only a single purpose.[1] At the same time that the nation has this bewildering and overlapping network of local and functional governments, it also relies to a great extent upon the federal government, whose direct general expenditures exceed those of all state and local governments put together.[2]

---

[1] Some professions in some states are also governed by organizations that are ostensibly private, but nonetheless have the legal authority to regulate a particular profession.

In some areas the federal government is over time playing a relatively larger role vis-à-vis state and local governments. The situation is, however, not simply or mainly one of federal activity supplanting decaying state and local governments. New local governments and functional authorities are being created, and since 1948 the federal government's share of total government expenditures (under any one of six alternative definitions) and of total government employees has declined, especially in relation to that of local governments.[3]

The network of grants or subsidies from higher levels of government to lower makes the situation even more complex. By one count, there are over 400 different federal grant-in-aid programs. Such grants are also becoming more important. Since 1957, federal grants-in-aid to state and local governments have been increasing at a compound rate of more than 14 percent per year.

The labyrinthian, overlapping network of state, local, and functional governments, dependent to an increasing degree on federal subsidies, is widely criticized. It is considered a waste of resources and an obstacle to intelligent planning and coherent public policy. However much their recommendations differ, all schools of thought seem to agree that the present pattern of governments is duplicative and unsatisfactory.

What principles ought to guide the development of a rational pattern of jurisdictional responsibility? Is a large-scale, centralized government a necessary condition of efficiency? Or a systematic reliance on small, local governments with rational boundaries? Is there any general or theoretical case to be made for subsidies from central to subordinate governments?

II

There is one issue that must be settled before we can proceed, for if it is not understood, the search for an optimal pattern of jurisdictional responsibility may seem trivial. This issue has been raised by those writers who emphasize that free bargaining among the parties concerned can deal with externalities in a Pareto optimal way. They rightly point out that the parties affected by an externality will always have an incentive to continue bargaining until all mutually advantageous bargains have been struck, at which time no one can be made better off without someone else being made worse off, so that there is Pareto optimality.

Many of those who like to emphasize this sort of argument are careful not to press it to an extreme and are quick to bring in bargaining costs or other considerations that limit the applicability of their argument. In this they are wise, for the "bargaining will continue until Pareto optimality is achieved" assumption could prove the efficiency, not only of laissez faire, but of anarchy as well. If free bargaining would always lead to Pareto optimality, all desirable public goods could be provided through voluntary action. Since violence is never Pareto optimal (unless it be between a sadist and a masochist), anarchy would be peaceful. The distributional implications of an anarchic but Pareto optimal society would not be appealing,[4] but then neither might those that free bargaining would lead to in a laissez faire economy.

What has just been said is not altogether frivolous, as is evident when we think of the implications of the "bargaining brings Pareto optimality" assumption for our search for principles of governmental structure. If, for example, there were no central government, there would be no loss of efficiency under the assumption of utopian bargaining, since local governments could through bargaining make national arrangements to take care of any uniquely national problems. Similarly, ideal logrolling in the national legislature could deal optimally with any local problem even in the absence of local government.

---

[2] George F. Break, "Changing Roles of Different Levels of Government," Table 1, p. 3 (forthcoming).

[3] Break, *op. cit.*; U.S. Department of Commerce, *Congress and the Nation*, p. 1393; and Robert A. Dahl, *Pluralist Democracy in the United States* (Rand McNally, 1967), p. 177.

[4] Those who could coerce others would enjoy extortionate gains. The anarchy might also be unstable and could degenerate into tyranny.

Another reason why the probably frequent failure of bargaining must be emphasized here is that where problems of government are at issue, the number of interested parties often numbers in the thousands or even millions. These large numbers not only raise what might be called the "transactions" cost of bargaining but also pose an additional problem. A large group can bargain with others only if it is organized, so that it can have representatives. But such an organization is itself a public good to the group it would help, and won't exist unless it enjoys the power of coercion or has some other way of giving its members an incentive on an individual basis, to support the organization.[5] Thus in many contexts where governmental action is at issue there may be no real possibility of a "bargaining" solution, since the groups among whom bargains could be struck aren't organized.

III

The subsequent analysis will be furthered by some special assumptions, many of which, as we shall see, can later be dropped. The first assumption is that the governments, authorities, and other institutions we consider produce only collective or public goods, which are defined for this purpose as goods such that it is not feasible to exclude nonpurchasers from their consumption. A second assumption is that every collective good affects some clearly delineated group or area, which can be as large as the entire population of the earth, or as small as the population of the smallest community. A breakthrough in pure science could benefit the entire world, and a program to combat air pollution could help only a particular community. Both of these examples satisfy our assumption rather nicely, but some other governmental services that are sometimes called public goods would not fit this assumption well, and we shall later have to relax it. A third initial assumption is that there is no complementarity in production among different public goods, in the sense that a government that provides one collective good cannot therefore provide a second collective good at lower costs than a unifunctional government could have. This assumption is probably also unrealistic in certain cases and will later have to be relaxed.

The argument that will be developed with the aid of these and other assumptions would better be described as a "premodel" than a "model," since its purpose is to set out a rudimentary and hopefully fundamental approach that should inspire the construction of elaborate and relatively realistic models based on less restrictive assumptions. There is perhaps some basis for the hope that this approach has promise in the fact that two other economists, Jerome Rothenberg and Gordon Tullock, have independently set out the essentials of arguments which, however different in some respects, in some ways parallel my own.[6]

---

[5] This point is demonstrated in my *Logic of Collective Action* (Harvard Univ. Press, 1965).

[6] See the forthcoming Universities-National Bureau of Economic Research volume on *The Economics of Public Output*, where Rothenberg, Tullock, and I discussed our approaches to this problem. See also my

The approach to be set out focuses on the problem of allocative efficiency rather than questions of stabilization and income redistribution, principally because stabilization and redistribution are obviously most appropriate to central governments. We ask here, "What are the necessary conditions for the allocative efficiency in the provision of collective goods?"

IV

In every case, one or more of the following three logically possible relationships between the "boundaries" of a collective good and the boundaries of the government that provides it will apply: (1) the collective good reaches beyond the boundaries of the government that provides it; (2) the collective good reaches only a part of the constituency that provides it; or (3) the boundaries of the collective good are the same as those of the jurisdiction that provides it.[7]

In the first type of situation, there is an obvious cause of inefficiency. The fact that the benefits of a given government's activity reach beyond its borders means that it provides an external economy, and when any maximizing unit provides an external economy, it tends to carry on its activity at a less than Pareto optimal level. In other words, situations of this kind leave a society with a smaller supply of the collective goods in question than would be optimal. There is some evidence that this actually happens in the case of military alliances, where one nation's military expenditures are an external economy to its allies, and in city-center governments which provide services to the whole of a metropolitan area.[8]

Now let us turn to the second possible type of situation, in which a collective good reaches only a subset of the population in a jurisdiction. Here there is what might be called an "internality," which normally has the same effect as the externality just described. To see this type of case in its most obvious form, suppose that the United States had a unitary government with no state or local jurisdictions. There would still be a need for local collective goods, such as air pollution controls in particular metropolitan areas.

In a situation of this type and a democratic political system with voting by majority rule, the provision of a collective good for a local area will hurt more people than it helps, even if Pareto optimality would have required that the collective good be provided. If the benefits are local and the taxes national, even a collective good which brings gains much greater than its costs will still create more losers than gainers. It is always possible to set up a special local taxing authority, but this will be at least a *de facto* government.

If all mutually advantageous bargains were struck, logrolling would insure that all collective goods that it was Pareto optimal to provide would be provided. But we have already argued that, especially where large groups of people are at issue, it will very often be the case that logrolling will not happen and that there will not be a Pareto optimal supply of public goods.[9]

---

discussion of papers on "Efficiency in the Government Sector," in *A.E.R.*, May, 1964, pp. 250–51, and Gordon Tullock's article on "Problems of Scale" (forthcoming in *Public Choice*).

[7] Possibilities (1) and (2) could both hold at the same time. It will soon be clear that this is particularly likely to lead to inefficiency.

[8] See Mancur Olson and Richard Zeckhauser, "An Economic Theory of Alliances," *Rev. of Econ. and Statis.*, Aug., 1966, pp. 266–79; and unpublished research on central city and suburban public expenditure patterns by Mr. Harvey Botwin, of Claremont College.

[9] In the United States Congress, logrolling probably leads to a greater than optimal expenditure when projects of a pork barrel type are at issue. In

Even when logrolling does occur, there is a likelihood that it will not provide a Pareto optimal supply of public goods. This is because any sharing of the marginal costs of a bundle of Pareto optimal projects which is not exactly proportional to the sharing of the gains at the margin will lead to a less than Pareto optimal expenditure. This, in turn, is because, under any other sharing of the costs, some members or members of the coalition will be paying more than their share and will find that marginal costs to them come to exceed their marginal benefits before a Pareto optimal amount has been provided for the society. They will then have an incentive to hold back. (This point was evident long ago from Lindahl's theory of voluntary public exchange.)[10]

There is one special case in which benefit boundaries smaller than jurisdictional boundaries will not lead to less than a Pareto optimal level of public expenditure. If x percent of the voters in a constituency are required to vote affirmatively before a measure is declared passed and a given collective good benefits more than x percent of the voters, there will presumably be a tendency for provision of a supra-optimal level of this good. This is because (as Gordon Tullock has shown) the minority that does not benefit from the good also pays taxes, and those who do enjoy it can vote for a level of provision that equates only their own tax burden at the margin with the total benefits of the collective good.

If there is this exception to the generalization that a failure of government boundaries to match the boundaries of collective goods leads to a less than Pareto optimal supply of collective goods, there is none to the rule that there are systematic forces which work against allocative efficiency in any situation where the boundaries of a government and a collective good it provides do not coincide. We must, then, argue provisionally (the argument will be modified somewhat later) that there is a need for a separate governmental institution for every collective good with a unique boundary, so that there can be a match between those who receive the benefits of a collective good and those who pay for it. This match we define as "fiscal equivalence."

Though it is much too early to make policy recommendations, it is already evident that both the "centralizing" and "decentralizing" ideologies are wrong, or at any event entail inefficiency. Only if there are several levels of government and a large number of governments can immense disparities between the boundaries of jurisdictions and the boundaries of collective goods be avoided. There is a case for every type of institution from the international organization to the smallest local government. It is a merit of the present approach that it can help explain the need for both centralized and decentralized government in the same framework.

---

most of these cases the projects are of a tangible, if not monumental, type, and a congressman is more likely to be identified in his district with such a project than with a general tax increase, which could not in any case usually be traced to any one package of local projects. The logrolled projects, moreover, normally fall under the jurisdiction of only one committee in each house, and this is often necessary for logrolling. The difficulty of logrolling across committee lines is illustrated by the fact that legislation which makes "sense" only if it applies only to a small number of areas, often can pass only after most congressional districts are made eligible for its direct benefits. If logrolling could readily occur with legislative proposals involving a number of committees, there would presumably be no such tendency. I am thankful to Charles Schultze for calling the difficulties of logrolling across committee lines to my attention.

[10] If bargaining proceeds solely in all-or-nothing terms, and marginal increments are never considered separately, the situation becomes more complex. But there would presumably still be some tendency in the same direction.

It is also already evident that some of the complaints about the proliferation of governments and the overlapping boundaries of different types of governments need to be greeted with skepticism. Though there are undoubtedly some redundant governments and some that are too small to serve their purposes properly, there is also a need for new governments as well, including some small ones. The metropolitan transport commissions, the port authorities, the Soil Conservation Districts, the suburban towns, and the neighborhood school boards that have been created are in many cases probably a response to the forces described in this paper. The forces we have described presumably have meant not only that expenditures on some public problems were at a less than Pareto optimal level, but also that some public problems weren't dealt with at all. This, in turn, must have led to the creation of new authorities or governments.

V

Patterns of residential segregation, such as are analyzed in Thomas Schelling's paper at this session, can also justify additional governments, if (and only if) it is assumed that integrated housing areas are not now a feasible (or desirable) alternative. Since different racial and ethnic groups often have different cultural backgrounds and tastes, they may want different types of collective goods. In cases where the sense of ethnic identity is very strong or where there is antagonism among different social groups, this is particularly important. People may then care passionately, not only about the level of expenditures on policemen, teachers, and other public authorities, but also about whether there is a civilian police review board or courses in Afro-American history and even about the race or religion of public officials at every level. The characteristics of the collective goods may affect not only the subjective evaluation of the services provided, but also their objective effects, as is evident whenever differences in language, accent, or social background impede communication between officials and citizens. The huge role of ethnic factors in American political history, especially at the level of city government, and the tense state of police-community relations in many Negro ghettos today, suggest that ethnic and religious factors can have a decisive impact on the nature of the collective goods that are demanded.

Though this question is much too complex to be settled here, it is clear that there is a case for a separate governmental institution for each ghetto, whenever group differences in taste for collective goods are important and the patterns of segregation in housing are not amenable to change. The collective goods then have relatively well-defined boundaries given by the boundaries of the ghetto, and governmental institutions which conform to these boundaries are a necessary condition for Pareto optimal provision of public services.

There is no assertion here that patterns of segregation in housing should or must be maintained, and no denial of the possible public interest in common arrangements for education and law enforcement which might ultimately lessen cultural and educational differences among ethnic and religious groups. Where education of children is at issue, the desire to satisfy consumer (parent?) preferences may be subordinated to a more general interest in the future strength and productivity of the nation. For these and other reasons, no general policy conclusion can be derived here. But the possibility of cases where matching boundaries of governments and ghettos are a necessary condition for fiscal equivalence and economic efficiency cannot be denied.[11]

---

[11] I have elaborated this part of my argument considerably in response to the most helpful comment of Daniel Ellsberg

VI

We have so far neglected economies and diseconomies of scale in the production of collective goods. We must now ask what influence the relationship of scale of government and production costs might have on the optimal mix of governmental institutions.

There will obviously be no modification of the principle of fiscal equivalence when the governmental unit with boundaries that match those of the collective good it provides happens to be of just the right size to produce the collective good at the lowest point on the average cost curve. Neither is there any objection to fiscal equivalence when a government whose boundaries are determined by this principle produces its collective good under conditions of decreasing cost. The fact that the collective good could be provided to a larger group at lower costs per person does not argue against fiscal equivalence because when there is such equivalence all of those who would benefit from the relevant collective good are already in the jurisdiction that provides it. A community building a levee for flood protection, for example, might wish it had more residents who could share the high fixed cost, but there could still be no social purpose in enlarging the boundaries of the government beyond the flood plain for which the levee has relevance.

The only case where the relationship between the costs of the collective good and the size of the clientele served would complicate the argument is that in which a government big enough to comprehend all of the benefits of its activity would have higher unit costs of production for the collective good than a smaller government. Here we have the only case where economies and diseconomies of scale in production play an independent role.

An example of this sort of situation might be education. Suppose it is true that a national school system would be so bureaucratic and cumbersome that it would have higher costs per student educated than a system with thousands of independent school boards. Even granted this supposition, there still appears to be some case for federal government involvement in education, since some of the benefits of education spread throughout the land (and, with "brain drains," even abroad). If children in one locality get a poor education, this could make the national democracy ultimately work less well. If the children of that locality often migrate to another locality, this poor education becomes a problem for the recipient community. Thus each local school district finds that its educational expenditures provide an external economy to the rest of the nation, and it therefore will spend too little on education.

What does fiscal equivalence require when the desire to internalize externalities calls for central government provision, and diseconomies of large-scale operation call for local provision?

The necessary condition for Pareto optimality is then local governments of a size that minimize unit costs,[12] and central government grants to these local governments that are just large enough to compensate the local government for the external benefits of its expenditures. The grants to the local governments must involve the sharing of the marginal costs of each additional unit of service in the same proportion as the benefits of that additional unit are shared between the locality and the nation at large. If a local government activity provides no benefits beyond its borders, it should get no federal subsidy (except for redistributive purposes, and this should be in the form of a lump-sum or block grant rather than one in which marginal costs are shared). If a local government produces something from which it gets no special benefit, the federal government should pay the full costs, as it does with other contractors who produce goods needed for national purposes.

---

[12] Unit costs may be a function, among other things, of the spatial distribution of the population and the level of per capita income in each area, both of which may be affected in turn by the level of taxes and public services in different jurisdictions. A formulation as simple as that in the text is satisfactory only if the possible interdependence between decisions about government boundaries and the provision and financing of collective goods, on the one hand, and the spatial distribution of population and economic activity, on the other, is ignored. I am indebted to Robert Hartman for helpful discussions of these issues.

There is, then, sometimes a case in principle for federal subsidies to state and local governments. In a period when improved transportation and communication and increased geographic mobility probably cause the external benefits of some local government activities to rise, we should not be surprised to see increasing federal subsidies to local governments. Nor should we assume that the great increases in subsidies of this type indicate that the local governments are obsolete: they could mean they are the most efficient purveyors of some types of collective goods.

VII

If some of the assumptions of the argument are relaxed, the situation becomes complicated. There is ultimately a need for an elaborate model that explores the implications of a wide variety of assumptions. But a rough idea of the effect of relaxing particular assumptions is easily gained.

It is clear that a relaxation of the assumption that there is no complementarity in the processes by which collective goods are produced tends to weaken the presumption that there should be a separate government for every collective good with a unique boundary. Some boundary discrepancies would be justified if collective goods could be provided more cheaply by governments that provided two or more such goods.[13]

The assumption that the boundaries of a collective good are given also needs to be relaxed, if the full range of what are usually called "public goods" is to be considered. Some of the goods that governments characteristically provide are such that it is sometimes practical to exclude nonpurchasers from their consumption. These are usually goods which can be provided either by the private or the public sector. Fire-fighting is an example of such a good. Though a fire in one building can spread to another, thereby preventing exclusion, it is also true that a fire department can sometimes let a particular building burn down without loss to anyone but its owner. With partly (or wholly) private goods it is possible to change the boundaries within which the good is provided. A fire-fighting district, for example, can be given a wide variety of boundaries.

What normative precepts would be relevant in such cases? The assumptions would have to be spelled out in greater detail to give a fully satisfactory answer. Where a government supplies a wholly private good, the logic of the argument here would demand that the provision of the good be left to the private sector. Where there is a quasi-public good that can feasibly be provided to areas with diverse boundaries, there is a good case for having the boundaries of the jurisdiction determined by the consideration of production cost. The size of jurisdiction and its particular boundaries would be set in such a way as to minimize the total cost of producing the desired good. Such external benefits as then spilled beyond the boundaries so determined should be covered by subsidies from a higher or more inclusive level of government.

---

[13] One way in which a multifunctional government may be able to cut costs is by reducing the need for legislators, officials, and elections, and saving voters the trouble of keeping informed about elected officials.

VIII

It is rarely appropriate to go directly from theory to policy. Reality is almost always too complex to permit policy recommendations derived solely from a single model, much less a premodel of the sort adumbrated here. The argument to the effect that governments that were either too small or too large would tend to provide a less than Pareto optimal supply of public goods might be countervailed, or even more than offset, by other considerations in any particular case. For example, as James Buchanan has shown, a progressive tax structure such that the median voter in a country pays less than the average rate of tax could lead to a more than optimal level of public expenditure. Countless other considerations might also be relevant in any given case.

Nonetheless, it seems highly likely that a great discrepancy between the boundaries of a collective good and those of the jurisdiction that provides it leads to problems of the kind described in this paper. This, in turn, argues that we can be reasonably certain that a broad array of governmental institutions is a necessary condition of Pareto optimal provision of collective goods, and that neither the extreme centralist nor the extreme decentralist position makes sense. It is also likely that many of the thousands of government and hundreds of intergovernmental subsidies that have been established in this country to deal with particular problems have in fact been needed, notwithstanding the complaints about the duplication and complexity in our system of government. The existing network of governments and subsidies has many serious faults, but is nonetheless probably better than most of the diverse arrangements that have been proposed to replace it.

MODELS OF SEGREGATION

By THOMAS C. SCHELLING
Harvard University

People get separated along different lines and in different ways. There is segregation by sex, age, income, language, color, taste, comparative advantage, and the accidents of historical location. Some segregation is organized; some is economically determined; some results from specialized communication systems; and some results from the interplay of individual choices that discriminate. This paper is about the segregation that can result from discriminatory individual choices.

My ultimate concern of course is segregation by color in the United States; but at the level of abstraction of this paper, any twofold distinction could constitute an interpretation—whites and blacks, boys and girls, officers and enlisted men. The only requirement in my model is that the distinction be twofold, exhaustive, and recognizable.

Skin color, of course, is neither dichotomous nor even unidimensional; but by convention the distinction is twofold, even in the U.S. census.

Economists are familiar with systems that lead to aggregate results that the individual neither intends nor needs to be aware of, the results sometimes having no recognizable counterpart at the level of the individual. The creation of money by a commercial banking system is one; the way that savings decisions cause depressions or inflations is another. In some cases small incentives can lead to striking results; Gresham's Law is a good example. My conjecture is that the interplay of individual choices, where unorganized segregation is concerned, is a complex system with collective results that bear no close relation to individual intent.

For some purposes an "unseen hand" of comparative advantage may sort people in a way that, though foreseen and intended by no one, corresponds to some socially efficient satisfaction of individual preferences. But we know a good many macrophenomena, like depression and inflation, that do not reflect any universal desire for lower incomes or higher prices. The worth of a new turnpike depends on constraining traffic below the density that would equalize its attractiveness with alternative routes. Typewriter keyboards, the pitches of screws, and left-hand automobile drive can be self-perpetuating in spite of inefficiency until an organized effort brings about concerted change.

A special reason for doubting any social efficiency in aggregate segregation is that the range of choice is so meager. The demographic map of almost any American metropolitan area suggests that it is easy to find residential areas that are all white or nearly so and areas that are all black or nearly so but hard to find localities in which neither whites nor nonwhites are more than, say, three-quarters of the total. And, comparing decennial maps, it is nearly impossible to find an area that, if integrated within that range, will remain integrated long enough for a man to get his house paid for or his children through school. The distribution is so U-shaped that it is virtually a choice of two extremes.

Some aspects of segregation lend themselves to quantitative analysis. Counting blacks and whites in a residential block or on a baseball team will not tell how they get along, but it tells something, especially if numbers and ratios matter to the people who are moving in or out of the block or being recruited for the team. And with quantitative analysis there are usually a few logical constraints, somewhat analogous to the balance-sheet identities in economics. Being logical constraints, they contain no news unless one just never thought of them before.

The simplest constraint is that, within a given set of boundaries, not both groups (colors, sexes) can enjoy numerical superiority. Within the population as a whole, the numerical ratio is determined at any given time; locally, in a city or a neighborhood, a church or a school, either blacks or whites can be a majority. But if each insists on being a local majority, there is only one mixture that will do it: complete segregation.

Relaxing the condition, if whites want to be at least three-fourths and blacks at least one-third, it won't work. If whites want to be at least two-thirds and blacks no fewer than one-fifth, there is a small range of mixtures that meet the conditions; and not everybody can be in the mixtures if the aggregate ratio is outside the range.

Other constraints have to do with small numbers. A classroom can be mixed but the teacher is one color; mixed marriages can occur only in the ratio of one-to-one; a three-man team cannot represent both colors equally, and even in a two-man team each member has company exclusively of one color.

In spatial arrangements, like a neighborhood or a hospital ward, everybody is next to somebody. A neighborhood may be 10 percent black or white; but if you have a neighbor on either side, the minimum nonzero percentage of neighbors of opposite color is fifty. If people draw their boundaries differently, we can have everybody in a minority: at dinner, with men and women seated alternately, everyone is outnumbered two to one locally by the opposite sex but can join a three-fifths majority if he extends his horizon to the next person on either side. If blacks occupy a sixth of the beds in a hospital and there are four beds to a room, at least 40 percent of the whites will be in all-white rooms.

There are several mechanisms by which blacks and whites, or boys and girls, can become segregated through individual choice. Whites may prefer to be among whites and blacks among blacks; whites may merely avoid or escape blacks and blacks avoid or escape whites; whites may prefer the company of whites, while the blacks don't care; and if whites can afford to live or to eat or to belong where the blacks cannot afford to follow, separation can occur.

Whites and blacks may not mind each other's presence, even prefer some integration, but, if there is a limit to how small a minority either color is willing to be, initial mixtures more extreme than that will lose their minority members and become all of one color; those who leave may move to where they constitute a majority, increasing the majority there and causing the other color to evacuate.

Evidently if there are any limits to the minority status that either color can tolerate and if initially complete segregation obtains, no individual will move to an area dominated by the other color. Complete segregation is then a stable equilibrium. The concerted movement of blacks into a white area or whites into black could achieve some minimum percentage; but in the absence of concert, somebody has to move first, and nobody will.

Let's examine a few of these mechanisms. Imagine a line along which blacks and whites (or men and women or Catholics and Protestants) have been distributed in equal numbers and random order, as in the line of plusses and zeros shown below. We expect them to be evenly distributed in the large but not in the small. If the colors or sexes or religions represented by plusses and zeros are content to live together in a ratio of about fifty-fifty, each finds himself in a satisfactorily mixed neighborhood if he defines his neighborhood as a long stretch of this line. If instead everybody defines his neighborhood as his own house and the neighbors on either side, a quarter of the whites and a quarter of the blacks are going to be surrounded by neighbors of opposite color. Satisfaction depends on how far one's "neighborhood" extends. For illustration, define everybody's "neighborhood" as extending four neighbors on either side, and suppose that everyone is content if half his "neighbors" are the same color as he. If fewer than half are his color, he moves in either direction to the nearest point (measured in the number he passes on the way) at which half his eight nearest neighbors are the same color as he. The particular row of plusses and zeros shown here corresponds to odd and even numbering in a column of random digits.

$$0\dot{+}000\dot{+}\ \ \dot{+}0+\dot{0}0\dot{+}\ \ +\dot{0}\dot{0}+\ \ +\ \ +\dot{0}+\ \ +\dot{0}+\ \ +\dot{0}0+\ \ +00+\ \ +0\dot{0}+\ \ +0\dot{+}\dot{0}+\dot{0}\dot{0}$$

$$+\ \ +\ \ +\dot{0}+\ \ \dot{+}00000\dot{+}\ \ \dot{+}\ \ \dot{+}000\dot{+}00\dot{+}\ \ +\dot{0}+\dot{0}+\ \ \dot{+}\dot{0}$$

I have put a dot over the individuals that are dissatisfied. It turns out that, of 35 plusses and 35 zeros, 11 plusses and 13 zeros are motivated to move. Two things happen as they move. Some who were content become discontent, as like neighbors move away or unlike ones move near; and some who were discontent become satisfied as like neighbors move near or unlike ones move away. Suppose that the dotted individuals move in turn, starting from the left, if they are still discontent when their turns come: rearrange the plusses and zeros by moving each dotted one to the nearest point where, inserting itself between two others, at least four of its eight neighbors are of its own color. This gives us the rearranged line:

$$00000000+\ \ +\ \ +\ \ +\dot{0}+\ \ +\ \ +\ \ +\ \ +\ \ +\ \ +\ \ +\ \ +\dot{0}\dot{0}00\dot{+}\ \ \dot{+}000\dot{+}0+\dot{0}+\ \ +\ \ +\dot{0}$$

$$+\ +\ +\ +\ +\ +\ +\ +\ +0000000000000000+\ +\ +\ +\ +\ +$$

Some who were going to move did not move after all. Eight have become newly discontent. We give them their turn and get this rearranged line:

$$00000000+\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +000000000$$

$$+\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +\ +0000000000000000+\ +\ +\ +\ +\ +$$

We end up with six groups of alternating color. Nearly half (thirty) have no neighbors of opposite color within four houses of them. Since we don't allow vacant spots, somebody is at the boundary of every group and has neighbors of opposite color; but, not only is everybody in a local majority as he wished to be, but by the efforts of each to achieve bare majority status they have together achieved an average majority status of more than five to one! This is not mathematical necessity: clusters of five would satisfy everybody at his minimum demands, but the actual clusters average twelve. (Alternating plusses and zeros or alternating pairs would also meet everybody's demands.) If people, though not wanting to be in the minority, prefer mixed neighborhoods, only forty of the seventy achieved it. Furthermore, anyone who wants some neighbors of opposite color, but not more than half, can move nearer the boundary of his cluster but will not move beyond; his movement will not change the clustering.

All of this is too abstract to be a motion picture of whites and blacks or boys and girls choosing houses on a road or even stools along a counter; but it is suggestive of some of the dynamics that could be present in individually motivated segregation.

Turn now to a different model. Suppose there is some area that both blacks and whites would prefer to occupy as long as the ratio of opposite color to one's own color does not exceed some limit. (This could be membership in an organization or occupation as well as a residential location.) We let this limit—call it "tolerance"—differ among the whites and also among the blacks.

Evidently the higher these limits, the more blacks and whites will be content to live in the area with each other. Evidently, too, the upper limits for the "most tolerant" whites and blacks must be compatible—their product must exceed one—or no contented mixture of any size is possible. And if nobody can tolerate extreme ratios, like 100 to 1, then, if the area is initially occupied by one color alone, none of the other would enter.

We can experiment with different distributions of tolerance to see what the process is by which the area becomes occupied by blacks or whites or a mixture, and to search for some principles that relate outcomes to the shapes of the curves, the initial positions, and the dynamics of entry and exit. There is no room for many alternatives in this paper, but the process can be illustrated.

What we are dealing with is a frequency distribution, separately for the whites and the blacks, of the upper limits to the ratios of opposite color to own color at which people will live in the area under consideration. The assumption is that anyone whose limiting ratio is exceeded by the prevailing mixture will go elsewhere. For ease of illustration suppose horizontal distributions from 2.0 to zero: for whites, the highest ratio of black to white that anybody can abide is two to one, the median white can tolerate a one-to-one ratio and the least tolerant cannot stand any blacks at all. On a diagram whose horizontal axis measures the white population and whose vertical axis measures the ratio of black to white on an arithmetic scale, the cumulative distribution will be a straight line intersecting the vertical axis at 2 and the horizontal axis at 100. For simplicity suppose the distribution of ratios of white to black that the blacks can tolerate is the same, and suppose that whites and blacks are equal in number.

To examine the dynamics we have to get whites and blacks on the same diagram. We translate the tolerance schedules into graphs expressing the absolute number of blacks whose presence can be tolerated by given numbers of whites, and vice versa. Keeping the whites ordered along the horizontal axis as they were in drawing up the frequency distribution—that is, with the most tolerant whites nearest the origin—we can plot, for a given number of whites, the maximum number of blacks whose presence they can tolerate. (We just multiply the number of whites by the ratio they can tolerate; the cumulative distribution of ratios translates into this absolute-number function exactly as a demand curve translates into a total revenue curve.)

The resulting two parabolas, Figure 1, divide the diagram into four regions. Any point beneath the inverted dish (the curve for whites, labeled *W*) is a point such that at least that many whites are satisfied with the presence of that many blacks: the whites present will not leave and additional whites will enter. Any point to the left of the blacks' curve (labeled *B*) represents a point at which at least that

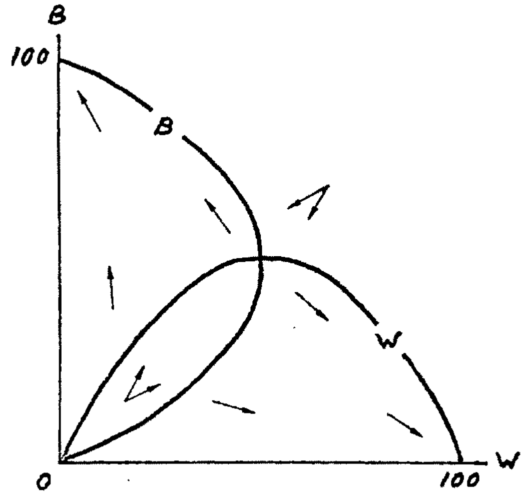


FIGURE 1

many blacks can tolerate the presence of that many whites: blacks will not leave and additional blacks will enter. Where the curves overlap the number of both blacks and whites present will be increasing; outside the curves, the numbers of both will be decreasing. Beneath the whites' curve but to the right of the blacks' curve, blacks will be evacuating and whites coming in; to the left of the blacks' curve but above the whites' curve, whites will be evacuating and blacks coming in.

There are two stable equilibria, one with exclusive occupation by blacks; the other with exclusive occupation by whites. The initial distribution of the two populations and the rates at which they move in or out will determine which one of the two colors eventually occupies and which one evacuates. Up to half of both colors could contentedly coexist at ratios near one to one, but the dynamics of entry prevent any mixture from stabilizing.

If the tolerance schedules are made steeper, the two parabolas can overlay each other as shown in Figure 2 (which corresponds to a slope of 5 and median tolerance of 2.5:1). There is now a stable mixed equilibrium. There are also stable equilibria at the two extremes. Again, which one would be obtained depends on initial conditions and rates of movement.

If whites outnumber blacks by two to one, the parabolas of Figure 2 will look as in Figure 3; the equilibrium mixture has disappeared. Whites numerically over-

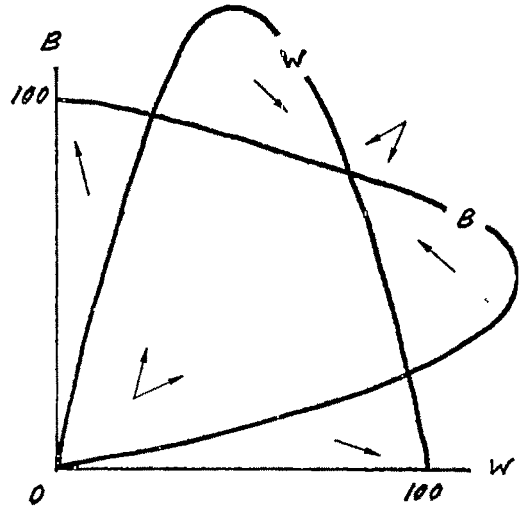


FIGURE 2

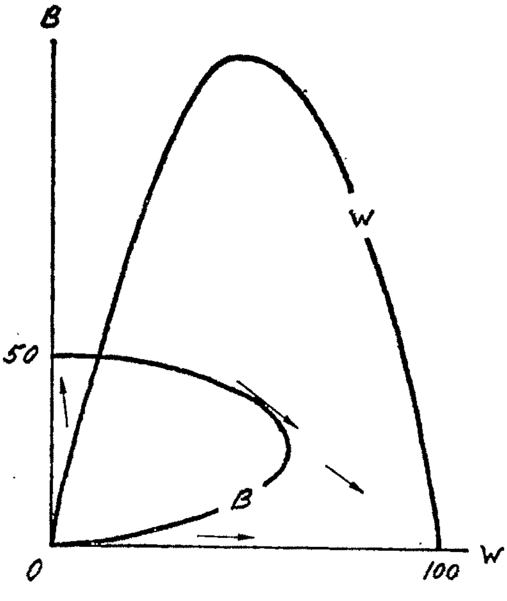


FIGURE 3

whelm the blacks, who evacuate (unless the initial mixture is in the thin upper-left slice). Limiting white entry can restore the stable mixed equilibrium. Interestingly, excluding some whites has the same effect as supposing the least tolerant whites to be more intolerant. Whether we limit whites in the area to half their total number or suppose that half the whites cannot tolerate any blacks at all, the tolerance schedule falls vertically at fifty whites, yielding intersecting curves (at thirty-six blacks) as in Figure 4.

This is but a small sample of possible results, using straight-line schedules and simple dynamics. There are no expectations in the model, no speculation, no concerted action, no restriction on the alternative localities available.

Just to mention two somewhat unexpected results: first, as we just saw, the polarized equilibria often come about because one color overwhelms the other; it is not the case, within the confines of this model, that the prospects for a stable mixed population are necessarily enhanced by an increase in the tolerance of one color for the other. (Make the least tolerant 60 percent of blacks and whites absolutely intolerant in Figure 1 and a stable equilibrium will occur at forty apiece.)

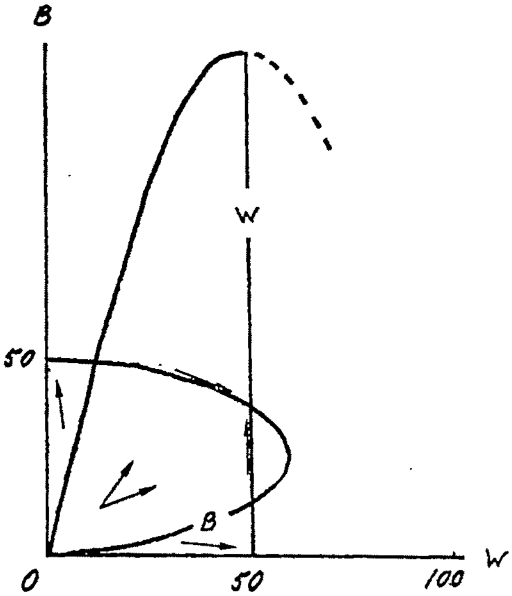


FIGURE 4

Second, the results do not depend on each color's having a preference for the absence of the other. We can equally suppose that most blacks and most whites prefer a color mixture, and reinterpret their tolerances as merely the upper limits to the ratios at which their preference for integrated residence is outweighed by numerical imbalance. The model fits both interpretations and produces the same results either way.

STRATEGIC INTERACTION AND RESOURCE ALLOCATION IN METROPOLITAN INTERGOVERNMENTAL RELATIONS

By JEROME ROTHENBERG
Massachusetts Institute of Technology

I. *Introduction*

This model concerns the influence that local government in a metropolitan area has on the distribution of economic activities within the area. It is one of a family of models the author has been studying to throw light on this issue. The basic situation here is that there does not exist a single metropolitan area government but a set of independent political jurisdictions. In the present model the area is assumed to be partitioned into a central city jurisdiction and a single suburban jurisdiction, both of constant area. Certain policy instruments controlled by these jurisdictions are assumed to influence the location decisions of households and businesses.

The model indicates how these instruments are used in a way to exert conflicting influences on such decisions, such that the respective opposing pulls are akin to a process of strategic interplay between two players in a variable sum strategic game. Since the model has not yet been subjected to complete analysis no attempt is made here to denote the exact form of the interplay or its convergence properties. But its source and essential character will be described.

The key actions are the location choices made by households (for residence) and business (for operation) between the central city (hereafter called "city") and the suburbs, and the responses to, and attempts to direct, these choices by city and suburban governments. We begin by motivating the private choices, and then indicate the aims of government in this matter.

II. *Residential Location Decisions*

Assume the metropolitan area contains a fixed population of N households, possessing a spectrum of income levels but similar tastes regarding urban and suburban living. Their preferences are a function of income. Suppose four classes of commodities enter their utility functions: a housing bundle (H), public sector services (G), privacy or spatial expansiveness (D), and a composite commodity (K) composed of everything else, for which relative prices remain unchanged. We assume that privacy depends on the nature of the individual housing package and of the neighborhood in which outside household activities must be carried on. Both of these depend on the density of land use. Since we assume that both urban and suburban space are respectively homogeneous, each can be characterized by a single density measure, and thus a single degree of privacy. We assume that within the relevant range more privacy is available in the suburb than in the city ($D_2>D_1$, where subscripts 1 and 2 throughout refer to the city and suburb, respectively).

Of the four utility arguments, H, G, D, and K, only K can be varied at will anywhere; the other three are spatially linked to a location decision: a household selects a package of H, G, and D either in the city or in the suburbs. Its choice of city or suburb is therefore a choice between different sets of commodity packages. Using residence in the city as vantage point, each household is assumed to calculate the complex of advantages which suburban living would possess over urban living, and then consider whether the cost of achieving these advantages warrants allocating the budget to the suburban option.

We must therefore introduce the prices of the four commodities: P_H, P_G, P_D, and P_K. We assume K is the numéraire (i.e., $P_K=1$). P_G is the price of one unit of public sector output. G is a single commodity for which quantity and quality variations can be expressed interchangeably in a single dimension. It represents an output level per household: e.g., a given level of police protection, transportation capability, fire protection, sanitation, health care, education quality, etc., per household. The resource capacity necessary to provide such a level for one household is available, of course, to provide the same level for the other households in the community as well (the public goods nature of public sector services). P_G then is the per capita resource cost of supplying this composite capability in the city or suburb.

P_D is more complicated. It is the price paid for the additional amount of privacy available in the suburb over the city. We derive it as follows. We suppose that business concentration in the city at the beginning is such that considerably more residents of the suburbs wish to be accessible to the city than city residents wish to be accessible to the suburbs. Thus city location has greater accessibility than suburban location. The cost of suburban privacy is inaccessibility. We assume for simplicity that accessibility to the core is the same everywhere in the suburbs. Let this common cost (in dollars) of suburban inaccessibility be V (in effect, a present capital value of transportation costs: travel outlays and the value of time). The cost of privacy is less than this.

Suppose $D_2=D_1$ and everyone were indifferent between urban and suburban living. Then the presence of inaccessibility cost V means that no suburban land would be occupied if the price of suburban land were less than urban land by less than V. Alternatively, if $r_1-r_2>V$ (where r_1, r_2 are per acre land prices in city and suburb, respectively), then households would increasingly migrate to the suburbs until the difference decreased to $r_1-r_2=V$. Thus, with no advantage from suburban living V would be completely offset by differential land prices. (A more realistic model would require increasing V with distance from the city, so that a metropolitan boundary could be set where the rural reservation price equaled the ever decreasing suburban offset price.)

If suburban location gives extra privacy, then households are willing to pay extra. Thus, migration proceeds until r_2 rises and r_1-r_2 falls below V, with the difference between them a measure of the cost of privacy. Thus,

$$P_D = V - (r_1 - r_2). \tag{1}$$

All four commodities are assumed to be superior goods, with privacy a luxury good (income elasticity of demand >1).

III. *Advantages and Costs of Suburban Living*

There are five types of differential advantages or costs associated with suburban living:

1. *Privacy.* The suburbs provide greater privacy than the city. The difference depends on the difference in density of land use between the two. This is assumed to be a function of the resident and business population in each. We assume not only a constant total household population ($N=N_1+N_2$) in the area but also a constant total business population ($B=B_1+B_2$). We can therefore express the amount of differential privacy provided in the suburbs as a function of suburban household and business population:

$$D=D(N_2, B_2). \tag{2}$$

Increases in either N_2 or B_2 decrease the differential privacy attainable, with

$$\left|\frac{\partial D}{\partial N_2}\right| < \left|\frac{\partial D}{\partial B_2}\right|$$

The price of this privacy has been given in (1).

An individual household i favors the privacy component of the suburban option if $U^i(D) > U^i[V-(r_1-r_2)]$ and a measure of the net privacy advantage for this household is:

$$\pi_1^i = \delta^i Q(h)\{U^i(D) - U^i[V - (r_1 - r_2)]\} \tag{3}$$

where

U^i is the cardinal (for simplicity) utility function of household i

π^i is net privacy advantage for household i

δ^i inverse of the marginal utility of income of household i

$Q(h)$ capitalization factor for computing present discounted values (h is interest rate)

$$= \sum_{w=0}^{\infty} \frac{1}{(1+h)^w}$$

The measure is given in discrete terms rather than in terms of D/P_D because the option is an all-or-nothing one (differing values of D cannot be chosen by individual choice).

2. *Welfare Redistribution.* Divide G into G_A and G_B, where G_A is the composite of public services which is equally available to all resident households, and not as a function of household income level; e.g., schools, parks, sanitation, police. G_B is the composite whose availability is strongly inversely related to income: call these welfare services in a broad sense. Assume that standards for welfare payments are fixed, so that the amount of total welfare services necessary in a community depends only on the size of the "welfare population"; i.e., on the distribution of households in the lower income tail. The suburbs will render an advantage to household i if the resources per household that must be redistributed through taxation to provide welfare services are less in the suburb than in the city—even if household i is a welfare recipient. This depends on the size and income distribution of suburban versus city population. We shall argue below that the specific attractions available in suburbs lead to a self-selection process in which suburbs come to possess a much smaller percentage of low-income households than the city. Moreover, this self-selection is augmented by explicit suburban public policy measures. Thus, welfare redistribution advantages will generally exist for each household i and may be substantial. We measure this as follows:

$$\pi_2^i = Q(h)\left\{ y_i\left[\frac{{}_1S_B(N_1,\ F_1(y))}{{}_1\bar{y}N_1}\right] - y_i\left[\frac{{}_2S_B(N_2,\ F_2(y))}{{}_2\bar{y}N_2}\right]\right\} \tag{4}$$

where

π^i_2 is the welfare redistribution advantage for household i

${}_1S_B$ is the resource cost of welfare services in the city

${}_2S_B$ is the resource cost of welfare services in the suburbs

$F_1(y)$ and $F_2(y)$ are distributions of income in city and suburb, respectively

y_i is household i's income level

${}_j\bar{y}$ is mean household income in jurisdiction j.

3. *Public Service Consensus.* The other component of G, G_A, is assumed to be a policy variable in each community and is decided by majority rule. We assume preferences are single peaked, so the majority position chosen is the median position. Since we have assumed G_A to be a superior good (described above as G), higher income households will want to buy more of it

than lower income households—if they are faced with the same prices. But prices are unequal, since public services are financed by taxes that are positive functions of income. We employ two types of taxes, one that may be interpreted as a tax on real estate improvements, the other a tax on land. Together they have roughly the effect of a proportional income tax for home owner-occupiers and at least a monotonic relationship for others. (Assume that the value of housing demanded and consumed is proportional to permanent income and that the latter equals current income. Then a proportional real property tax on land and improvements is equivalent to a proportional income tax.) Therefore,

$$G_A^D(y_i, P_{G_A}i) \lesseqgtr G_A^D(y_j, P_{G_A}j)$$

for $y_j > y_i$, where G_A^D is the demand for G_A, y_i the income level of household i, and $P_{G_A i}$ the relevant tax price per unit G_A to household i.

However, consider two communities, equally populated but with different mean incomes, the population of the richer having equal incomes, that of the poorer highly unequal incomes: ${}_2\bar{y} > {}_1\bar{y}$. For any G_A, the total resource cost is $S_A = S_A(G_A, N_j)$ where N_j is the population of each community. The proportional income tax rates are respectively,

$$t_1 = \frac{S_A(G_A, N_1)}{{}_1\bar{y}} \quad \text{and} \quad t_2 = \frac{S_A(G_A, N_2)}{{}_2\bar{y}N_2}$$

$t_2 < t_1$: the tax rate is lower in the richer community. Now, a household i with $y_i > {}_1\bar{y}$ will demand more G_A when in the homogeneous richer community 2 than in the poorer, more income dispersed community 1. A move from the latter to the former lowers the real price of G_A to household i—makes it better off (the marginal impact of the move on ${}_1\bar{y}$, ${}_2\bar{y}$ is assumed negligible). The choice of location reflects the welfare increase.

The level of G_A chosen by the community will depend on these same characteristics. If ${}_2\bar{y} > {}_1\bar{y}$ and $t_2 < t_1$ for each given G_A, then $G_A^D({}_1\bar{y}, t_1) < G_A^D({}_2\bar{y}, t_2)$: the mean income family in community 2 wants a larger G_A than does his counterpart in community 1. Assume a symmetrical income distribution in the latter (as we already have in the former). Then $\bar{y} = \tilde{y}$ in both ($\tilde{y}$ is the median income level). Assume further that price and income elasticities are nearly constant, especially over income levels near ${}_2\bar{y}$. Now, if everyone paid the same total tax, the array $G_A^D(y_i | P_{G_A 0})$ $i = 1, 2, \cdots, N$ would monotonically increase in income, and the median income household $(y_k = \tilde{y})$ would have the median demand for G_A: i.e., $\tilde{G}_A^D = G_A^D(\bar{y} | P_{G_A \bullet})$ and be group chosen. Now set the same income tax rate for all, so that the poor pay less, the rich more. Under our relative constancy of elasticities $\tilde{G}^D$ remains relatively unchanged, and $\tilde{G}_A^D \simeq G_A^D(\bar{y} | P_{G^S \bullet})$ (where $\simeq$ means approximately equal to) since as many poor with lower prices want more G_A than $G_A^D(\bar{y} | P_{G_A \bullet})$ as rich with higher prices want less G_A than $G_A^D(\bar{y} | P_{G_A \bullet})$. (With $\eta_y = 1 = \eta_p$, $G_{Ai}^D = \tilde{G}_A^D$ all i.) So ${}_1G_A = G_{A1} \simeq G_A^D({}_1\bar{y}, t_1)$ and ${}_2G_A = \tilde{G}_{A2} \simeq G_A^D({}_2\bar{y}, t_2)$; therefore, since ${}_2\bar{y} > {}_1\bar{y}$, ${}_2G_A > {}_1G_A$ (where ${}_jG_A$ is the level of G_A selected by community j): the richer community demands higher G_A.

We argue that the city and suburb resemble communities 1 and 2, respectively, the city poorer and with highly dispersed incomes, the suburbs richer and with considerably more homogeneous incomes.

Consider the consumer surplus gain for household i with $y_i = {}_2\bar{y} > {}_1\bar{y}$ if it should move from city to suburb. Price falls from some $P_{G_A \bullet}$ to $P_{G_A 1}$ and G_{Ai}^D rises from $G_A^D(y_i, P_{G_A \bullet})$ to $G_A^D(y_i, P_{G_A 1}) = G_A^D(y_i, P_{G_A 0}) + \Delta P_{G_A}\eta_p^i$, (where η_p^i is i's price elasticity of demand). The community-chosen G_A increases from $\tilde{G}_{A1}$ to $\tilde{G}_{A2}$, or $G^D({}_1\bar{y},\ P_{G_A1})$ to $G^D({}_2\bar{y},\ P_{G_A2})$, which is approximately $G^D({}_1\bar{y},\ P_{G_A1})+\Delta P_{G_A}\eta_p^k$ (where η_p^k is the price elasticity of demand for household k with income$={}_2\bar{y}$). The consumer surplus gain for household i is the conventional welfare triangle, and is the measure of the political service consensus impact on i:

$$(5)\qquad \pi_3^i \simeq 1/2Q(h)\left(\frac{t_1y_i - t_2y_i}{G_A}\right)\eta_p^k$$

$$= 1/2Q(h)y_i\left(\frac{t_1 - t_2}{G_A}\right)\eta_p^k$$

$$= 1/2Q(h)y_i\left(\frac{{}_2\bar{y} - {}_1\bar{y}}{{}_1\bar{y}{}_2\bar{y}}\right)\eta_p^k$$

(Assume ${}_2S_A=G_AN_2$ and ${}_1S_A=G_AN_1$. Then ${}_jS_A/N_j=G_A$. So $t_2=G_A/{}_2\bar{y}$ and $t_1=G_A/{}_1\bar{y}$; therefore,

$$t_1 - t_2 = \Bigg(\frac{G_A}{{}_1\bar{y}} - \frac{G_A}{{}_2\bar{y}} = G_A\left(\frac{{}_2\bar{y} - {}_1\bar{y}}{{}_1\bar{y}{}_2\bar{y}}\right).)$$

4. *Externality Induced Tax Burden in City.* As stated above, the initial concentration of commercial and industrial business in the city means that considerably more suburban residents will wish to be present in the city than city residents will wish to be present in the suburbs. The presence of outsiders in each jurisdiction, using its public and private facilities, will affect the resource cost of supplying every level of G_A (a quality level of output to its own residents) within that jurisdiction, either by affecting the quality of service through congestion or by requiring a given capacity to be divided over a larger affected population. The greater the outsider use the greater will be the resource cost of supplying each G_A. The notable asymmetry of cross-use between city and suburb means that these real externalities will on balance be a burden for the city government. (I have dealt extensively with this in "Local Decentralization and the Theory of Optimal Government," Universities-National Bureau Conference on Economics of Public Output, 1968.)

The financial burden will be lessened to the extent that the city government can tax suburbanites for these externality costs. But local governments are limited in the types of taxation they may effectively employ, either by law or the mobility of the tax base. Besides, the specific outsider uses that generate the externalities are extremely difficult to reach through user charges. Since most changes fall on residents as well, any given total will include higher revenues from residents. Thus, taxation of the suburbanite will typically fall short of the total externality cost incurred. So city residents will be forced to pay a unit price for G_A which is inflated by the amount of externality cost unrecouped from suburbanites. The unit price of G_A in the suburbs will not include any such component. Suburbanites who pay some charge for their presence in the city typically pay less than the social cost they incur. More relevantly, they receive public services in the city over and above what they obtain (and pay for) in the suburbs. For this they pay an amount which is demonstrably less than what city accessibility and its public services are worth (since they are willing to pay it). Their city tax liability is a lower bound of the worth of this package.

Let $E({}_1G_A,\ N_1,\ N_2)$ be the externality cost function, relating the added costs to the city-desired G_A and its client resident population, and the size of the potential outsider use population. Let $(1-a)$ be the proportion of these which is collected from suburbanites. Then the inflation of public good price which stems from externalities is an advantage of suburban over city location and is measured for household i as:

$$(6)\qquad \pi_4^i = Q(h)y_i\left[\frac{aE({}_1G_A,\ N_1,\ N_2)}{{}_1\bar{y}N_1}\right].$$

5. *Minimum Lot Size.* As we shall indicate below, suburban governments use residential zoning as an instrument of public policy. We assume this to be in the form of specifying the maximum number of households which can legally occupy each acre. Thus, it specifies the minimum size lot which a resident must use in consuming housing services. This has a utility cost for households which would have chosen a smaller lot as part of their optimal housing package, but not for those whose lot size was no smaller than the minimum. We assume no such zoning operates in the city. Therefore, this constraint is a negative advantage of suburban location, and is measured for household i by:

$$\begin{aligned} \pi_5^i &= \delta^i Q(h)[\hat{U}^i(\hat{K}, \hat{H}, \hat{D}, \hat{G}) \\ &\quad - U^{i*}(K^*, H^*, D^*, G^* \mid H_1 \geq H_1^{Z_R})] \\ &= \delta^i Q(h) H_1^{i*}(y_i, r_2) \end{aligned} \tag{7}$$

where $\hat{U}^i$ is the utility for i's optimum budget allocation $(\hat{K}, \hat{H}, \hat{D}, \hat{G})$, given its tastes, budget and relative prices, U^{i*} is U for the optimum allocation constrained by minimum lot size, H_1 is the land component of the housing bundle, $H_1^{Z_R}$ the minimum lot size set by zoning regulation Z_R.

IV. *Consequences of Locational Advantage*

The above are five respects in which suburban and city location differ. It is important to examine the consequences of these differences. First is the question of who benefits; second, the question of how these benefits vary over different sized suburban populations.

Examination of the five impacts shows that benefits apply primarily to the richer portion of N. π_5^i is an actual disadvantage for low-income groups, the lower the income group the larger the disadvantage. For those with high enough income to wish to use a lot larger than the legal minimum it is no disadvantage at all. The first advantage favors the rich. Their utility valuation of extra privacy is high while that of the money costs of accessibility is relatively low. The opposite is true for lower-income groups. The second and fourth are proportionally higher the higher household i's income level. The third is not only proportionally higher on this score, it is additionally higher the greater the difference between mean income in the suburb and the city.

The impact of this is that the suburbs will tend to attract chiefly the highest income classes (having assumed away demographic differences among households). The suburbs will tend to be a high income, highly homogeneous community relative to the city. It is clear that minimum lot zoning is a device that enhances the strength of the other suburban advantages.

The second issue, concerning how these benefits vary with suburban population size, is illuminated by this same factor. If suburban location is most attractive to the rich, we assume the sequence of migration will tend to be by decreasing household income level. Then the size of N_2 and the expected income level of the last migrant in the queue will be inversely related. Therefore the benefits for the last migrant will decline in every category with increasing N_2.

In the privacy impact (equation 3) the utility value of D declines because: (1) D itself declines with increasing N_2, since increasing density decreases suburban privacy (thus incidentally decreasing benefits for everyone earlier in the queue as well); (2) the utility significance of any given value of D falls because D is a luxury good. Moreover, the utility significance of the money costs of inaccessibility rises. For welfare redistribution (equation 4), a suburban case load begins to grow with greater income dispersion and the tax rate rises as mean income falls. So the differential burden between city and suburb decreases. Besides, the last migrant's income level is lower. In the consensus impact (equation 5) the difference in suburb-city mean income levels, as well as the last migrant's income, are lower. In the externality impact (equation 6), the lower last migrant's income has the same effect. Finally, the lot size constraint (equation 7) comes to be operative at some N_2 and increases thereafter with larger sizes as the last migrant's housing demand decreases with his income.

As N_2 increases, some benefits approach or become zero, and categories 1 and 5 actually become negative. Converting money magnitudes into utility terms for the appropriate last migrant, there will be some N_2 at which the net utility gain is zero for the last migrant. This household is the marginal migrant, households further in the queue having negative net gains. Thus, for given states of the policy variables to be introduced in the next section an equilibrium N_2 is determined: that size for which the last migrant has zero net gains. All households with positive net gains are assumed to suburbanize, all with negative net gains to locate in the city. A change in the policy variables will change this equilibrium.

V. *Public Policy Instruments and the Goals of City and Suburban Governments*

The city government is assumed to possess two policy instruments that can influence locational decisions: the percentage of externality costs which it will recover from suburbanites (a in equation 6), and the land tax rate (b_1). The improvement tax rate (which essentially carries the weight of an income tax) is a residual of expenditure decisions and these two tax policy decisions.

The suburban government is assumed to possess three relevant instruments: first, the land tax rate (b_2); second, a residential zoning power (Z_R), which specifies the minimum permissible sized lot per household (and thus, with our assumption of equal density, the maximum permissible number of households per acre); third, a business zoning power (Z_B), specifying the maximum permissible amount of business activity per suburban acreage (in terms of value of improvements: which we assume closely associated with size of value added).

While residential location is endogenous in the present model, business location is not. It is assumed that important suburban advantages exist for businesses but their locating there is limited by zoning restrictions. The amount of business activity is thus assumed to equal the maximum permitted by zoning. A change in this ceiling through Z_B will be exactly matched by business location. We further assume that suburban businesses do not generate external costs on the city (although a more complex model on which I am working in which business location is endogenous does allow this to increase the symmetry of outsider cross-use and so decrease net externalities), and that public services to businesses exactly match tax revenues from businesses (so that their location does not affect financial burdens on households).

The land tax and improvements tax are separately treated because they presumably have different effects on the location decision. The immobility of land (its inelastic supply) relative to that of improvements means that a tax on the latter will have much greater effect on location decisions than one on the former (often treated as zero effect). An increase in b_i implies that a larger part of total outlay is to be financed by locationally neutral means (except that, as we shall see, a land tax has indirect effects).

The goal of both the city and the suburban government is assumed to be to maximize the total value of land within its jurisdiction. Each in effect attempts to enhance its productivity as a site for economic activity. This goal is intended to represent the majority interests of each constituency, but this is obviously so only in a model where all households are owner-occupiers. To avoid spelling out a complicated theory of political representation we make the non-trivial assumption that they are.

We turn now to the political maximization problem and the statement of the complete model.

VI. *Land Value Maximization: the Complete Model*

A. *Goal:*

1. City: max $R_1 = M_1 r_1(N - N_2,\ B - B_2,\ r_2)$ with respect to a, b_1
2. Suburb: max $R_2 = M_2 r_2(N_2,\ B_2,\ r_1)$ with respect to b_2, Z_R, Z_B

B. *Behavior Equations*

1. $r_1 = r_1(N - N_2, B - B_2, r_2)$
2. $r_2 = r_2(N_2^q, B_2, r_1, Z_R)$
3. $N_2^q = \lambda[N, \pi(N_2)]$
4. $N_2 = N_2(N_2^q, Z_R, Z_B)$
5. $\pi_1(N_2) = \delta^{i(N_2)} Q(h)$

$$\cdot \left\{ U^{i(N_2)}[D(N_2^q,\ B_2,\ Z_R)] - U^{i(N_2)}(P_D) \right\}$$

$$= \delta^{i(N_2)} Q(h) \left\{ U^{i(N_2)}[D(N_2^q, B_2, Z_R)] - U^{i}[V(Z_B) - (r_1 - r_2)] \right\}$$

6. $\pi_2(N_2) = Q(h) y_{i(N_2)} \left[\dfrac{{}_1S_B(N_1,\ F_1(y))}{{}_1\bar{y} N_1} - \dfrac{{}_2S_B(N_2,\ F_2(y))}{{}_2\bar{y} N_2} \right]$
7. $\pi_3(N_2) \simeq 1/2 Q(h) y_{i(N_2)} \left(\dfrac{{}_2\bar{y} - {}_1\bar{y}}{{}_1\bar{y}\,{}_2\bar{y}} \right) \eta_p^k$
8. $\pi_4(N_2) = Q(h)\,{}_1G_A y_{i(N_2)} (T_1 - T_2)$
9. $\pi_5(N_2) = \delta^{i(N_2)} Q(h) [\hat{U}^{i(N_2)}(\hat{K}, \hat{H}, \hat{D}, \hat{G}_A)$

$$- U^{i(N_2)} K^*, H^*, D^*, G_A^* \,|\, H_1^* \geq H_1^{Z_R})]$$

$$= \delta^{i(N_2)} Q(h) H_1^{*i(N_2)}(y_{i(N_2)}, r_2, Z_R)$$

10. $D = D(N_2^q, B_2, Z_R)$
11. ${}_jG_A = G_A({}_jS_A, N_j, {}_j\bar{y}) \quad j = 1, 2$
12. $B_2 = B_2(B, Z_B)$
13. $T_1 = \dfrac{1}{N_1\,{}_1\bar{y}} [{}_1S_A({}_1G_A, N_1) + aE(G_1, N_1, N_2) + C(1 - a) - b_1 R_1]$
14. $T_2 = \dfrac{1}{N_2\,{}_2\bar{y}} [{}_2S_A({}_2G_A, N_2) - b_2 R_2]$
15. ${}_jS_B = S_B(N_j, F_j(y)) \quad j = 1, 2$
16. $r_j = (1 - b_j) g_j \quad j = 1, 2$
17. $V = V(Z_B)$
18. $F_J(y) = f_j(F(y), N_j) \quad j = 1, 2$
19. $y_{i(N_2)} = y(N_2)$
20. ${}_j\bar{y} = {}_j\bar{y}(N_j, F_j(y)) \quad j = 1, 2$
21. $H_1^{Z_R} = H_1(Z_R)$
22. ${}_jS_A = S_A({}_jG_A, N_j) \quad j = 1, 2$
23. $\delta^{i(N_2)} = \delta(y_{i(N_2)})$

C. *Definitions*

1. $R_j = r_j M_j$
2. $G = G_A + G_B$
3. $P_D = V - (r_1 - r_2)$
4. $\pi(N_2) = \sum_{i=1}^{5} \pi_i(N_2)$
5. ${}_j\bar{y} = \bar{y}(F_j(y))$

D. *Policy Variables:*

1. City: a, b_1
2. Suburb: b_2, Z_R, Z_B

E. *Predetermined Variables:* N, B, Q

F. *Definitions of Variables*

1. N, N_1, N_2 total, city, suburb population
2. B, B_1, B_2, total, city, suburb business activity
3. r_1, r_2 per acre price of city land, suburb land
4. $N_2^{\S}$ population desiring suburban residence
5. $\pi(N_2)$ total suburban net benefit function
6. $\pi_1(N_2)$ privacy net benefit function
7. $\pi_2(N_2)$ welfare distribution net benefit function
8. $\pi_3(N_2)$ consensus net benefit function
9. $\pi_4(N_2)$ externality net benefit function
10. $\pi_5(N_2)$ lot size loss function
11. D level of privacy
12. $\delta^{i(N_2)}$ inverse of the marginal utility of income of smallest total net gain migrant for given N_2
13. $Q(h) = \sum_{w=1}^{\infty} \frac{1}{(1+h)^w}$

    capitalization factor, h is interest rate
14. P_D price of privacy
15. V capital value of cost of suburban inaccessibility (travel costs)
16. $y_{i(N_2)}$ income level of smallest total net gain migrant for given N_2
17. ${}_jS_B$ cost of welfare load in community j
18. $F_j(y)$ income distribution in community j
19. ${}_j\bar{y}$, ${}_j\tilde{y}$ mean, median income in community j
20. ${}_jG_A$ per household level of government services other than welfare in community j
21. η_p^k price elasticity of demand for G_A by median income household
22. T_j improvement (income) tax rate in community j
23. $U^i(K, H, D, G_A)$ utility function of household i; composite commodity, housing, privacy, public services other than welfare; ^unconstrained optimum; *optimum constrained by minimum lot size
24. $C(1-a)$ collection cost function for taxing suburbanites
25. R_j total land value in community j
26. M_j total acreage in community j
27. ${}_jS_A$ resource cost of non-welfare public services in community j
28. g_j no-tax value of land per acre in community j
29. $H_1^{Z_R}$ minimum lot requirement
30. a percentage of externality cost not taxed from suburbanites
31. b_j land tax rate in community j
32. Z_R, Z_B residential, business zoning regulation

G. *Comments*

Space permits only the briefest additional comments on the model.

1. Land values: Land values depend on total activity demand for space. City and suburban values are related—they can neither diverge nor converge overly. Residential demand depends on prospective net benefits and zoning constraints (the size of the suburb-desiring population for which the smallest gainers would gain nothing versus the largest population permissible by residential and business zoning); business demand depends on the zoning ceiling.

2. Net benefits: They generally depend on the policy variables through an influence on the income distributions in any N_1, N_2. Resource cost of non-welfare public output can show first increasing then decreasing returns to scale with increasing population. Benefit equation types 1–4 show both benefits and costs. Type 4 has been enlarged to allow for scale economies and diseconomies, tax collection costs and the substitution of location-neutral land taxes for inprovement (income) taxes. Each $\pi_m(N_2)$ indicates the benefits to that household which has the smallest total net gain for each given N_2. Present value calculations assume unchanging expected benefits over time. Travel costs decrease with increasing business activity in the suburbs. Benefits are made comparable in

money terms by using marginal utilities of income.

VII. *Maximization Conditions*

We solve the maximization problem by deriving first and second order conditions in the usual way. We shall simply state the first order conditions.

A. *City*

1. $E=C'$

(1) (2)

(1) Externality costs
(2) Marginal suburbanite tax collection cost

2. $\delta^{i(N_2)}U^{i(N_2)'}g_1=\frac{{}_1G_Ay_{i(N_2)}M_1}{N_1\,{}_1\bar{y}}(r_1-b_1g_1)$

(1) (2)

(1) Accessibility cost impact
(2) Public service price impact

B. *Suburb*

1. $\delta^{i(N_2)}U^{i(N_2)'}g_2=-\frac{{}_1G_Ay_{i(N_2)}M_2}{N_2\,{}_2\bar{y}}(r_2-b_2g_2)$

(1) (2)

(1) Accessibility cost impact
(2) Public service price impact

If the population ceiling through zoning is inoperative:

2a. $U^{i(N_2)'}D_{ZB}=H_{1ZB}^{*i(N_2)}$

(1) (2)

(1) Privacy impact
(2) Minimum lot cost impact

If the population ceiling through zoning is operative:

2b. $-r_{2ZB}=\delta^{i(N_2)}Q(H)r_{2\lambda}\lambda_x[U^{i(N_2)'}D_{ZB}$
$-H_{1ZB}^{*i(N_2)}]$

(1) (2)

(1) Population ceiling impact
(2) Suburban location gains impact

3. $D_{B_2}B_{2ZB}=V_{ZB}$

(1) (2)

(1) Privacy impact
(2) Accessibility cost impact

VIII. *Intergovernmental Strategic Interplay*

The equations for each government, solved simultaneously, give each the optimal values of its policy instruments, its optimal population and business activity size, and its maximum total land value. (These optima do exist with non-corner solutions for at least some pairs of values.) But each such set relates to a particular set of values for the other government's policy instruments. As each changes its initial set toward optimal values it changes the circumstances for the other and requires a new adjustment. Thus we have a familiar reaction function problem. It is well beyond the scope of the present paper to examine the convergence properties of this system. But it is characterized by a play and interplay conflict between the two governments to attract households and businesses away from one another. In a further work I hope to spell out the implications of this interplay for the efficiency of resource allocation in metropolitan areas.

ROUND TABLE ON ALLOCATION OF RESOURCES IN LAW ENFORCEMENT

THOMAS C. SCHELLING, *Chairman*

ANTHONY DOWNS: In every big city, police officials must use some basic principles in order to allocate their limited forces in space (and time). In the past, these principles have often been implicit rather than openly expressed. But improved operating methods in major police departments have focused on the need to define objectives clearly. So it may be useful to spell out some possible alternative strategies.

Surprisingly, quite a few different objectives for police activity can be conceived of—and they lead to widely varying allocation strategies. Some of the possible goals of allocation are: (1) minimizing total index crime within the city as a whole; (2) providing an average police response to all calls for service within some minimum time (say, two minutes); (3) providing equal protection to all parts of the city—defining "protection" in terms of resource inputs; (4) providing equal protection to all parts of the city—defining "protection" in terms of safety outputs; (5) creating equal work loads on all operating units; and (6) developing an effective capability to respond quickly with the "massive" force needed to confront large-scale violence, such as riots.

Allocation strategies must be developed within the constraints imposed by certain widespread conditions affecting big-city police work. Five such constraints are especially significant. The first is the much higher incidence of reported crimes and certain other calls for service in low-income neighborhoods than in middle-income and higher-income areas. Thus, police protection is similar to education in one key way: the major determinant of the quality of service received in a given area is the nature or behavior of the consumer group residing there rather than the quality or quantity of public resource inputs made there. The second constraint is the concentration of calls for services in certain periods of the day (notably between 4:00 p.m. and midnight) and on certain days of the week in definite patterns. The third is the fact that police are normally concerned with a great many services in addition to combating crime. The most significant is regulating traffic, but many others—such as settling domestic quarrels—abound. The fourth is the much greater residential density in low-income areas than in high-income areas (except in high-rise apartment areas, as in New York). The last constraint is the limited resources police officials have with which to meet burgeoning demands for called for services.

Given these constraints, the six policy objectives stated initially can be translated into appropriate allocation strategies. Minimizing index crime would result in concentration of almost all police forces in the highest-crime areas during the busiest periods. The areas receiving such intense coverage would be low-income neighborhoods and large downtowns. This strategy might completely denude outlying higher-income areas of any police protection at all, since crime rates are vastly higher elsewhere (sometimes the ratio is as high as 35 to 1 for crimes against persons per 100,000 population). In contrast, providing a minimum average response-time in all parts of the city would cause significant dispersal of forces. Even so, a higher patrolman per capita ratio would still remain in low-income areas because they generate more calls for service. This strategy might be used as a constraint on others rather than as a dominant objective, particularly if the department can meet the minimum response-time standard without using all of its forces to do so.

Providing equal resource inputs to all parts of the city (that is, per 100,000 residents) would result in a highly unequal crime rate distribution. It would not minimize crime either, because high-income areas would be "overprotected." Moreover, work loads would also be quite unequal: officers assigned to low-income areas would slave away while those in wealthier areas loafed. In contrast, providing equal protection to all parts of the city as defined in terms of outputs—that is, crime rates—is probably impossible without enormous expansion of most big-city police forces. Crime rates—particularly for crimes against persons—are so much higher in the lowest-income areas than in the wealthiest areas that even putting all police in the former might not drive down rates there as low as in the latter (though those would certainly rise).

A strategy of equalizing work loads would closely resemble a combination of crime minimization and maintaining minimum response time all over the city. Like these, it would also involve concentrating police on the high-activity shift rather than having equal force levels on all three shifts. The last objective—maintaining effective capacity to respond quickly with massive force—requires holding a certain proportion of all resources in reserve. This in turn reduces "normal" performance capabilities throughout the city. During recent periods when such "special forces" were constantly being used in some big cities to cope with civil rights and other demonstrations, "normal" crime rates soared.

The above are "pure" strategies in the sense that each is designed to serve a single objective. In reality, all police departments have multiple objectives, which include generating sufficient political support throughout the community to maintain their incomes and the community's respect. So actual allocation strategies involve combinations of those described above. But every such combination includes provision of at least some minimum level of service to the middle-income and higher-income districts where crime is low but most taxpayers reside.

WILLIAM M. LANDES: A common belief about our legal system is that persons arrested for crimes will have their cases decided in a court trial. In reality, the outcome of a case rarely depends on a trial as most cases (probably more than 90 percent) are disposed of before trial through negotiations between the prosecutor and suspect, resulting in either a guilty plea or a decision not to charge the suspect.[1] In this note, I briefly outline a model that explains the conditions under which a pretrial settlement (PTS) or trial will take place, and then use the model to analyze the existing bail system, court delay and the effects of charging a money price for the courts. An expanded version of this model together with a detailed empirical analysis is being completed.

The Model. Assume the following: (1) the prosecutor and suspect take as given the sentence (S) the suspect would receive if convicted in a trial; (2) the probability of conviction is positively related to the prosecutor's resource input (R^*) and negatively related to the suspect's input (R); (3) the prosecutor behaves to maximize the expected number of convictions for n cases weighted by their respective sentences subject to a constraint on the total resources available to his office; and (4) the suspect maximizes the expected utility of his endowments or wealth in various states of the world.

The prosecutor maximizes

$$(1)\quad E(C) = \sum_{i=1}^{n} P^*_i S_i + \lambda \left(B_0 - \sum_{i=1}^{n} R^*_i \right)$$

where P^*_i is his estimate of the probability of conviction in the ith case ($i = 1, \cdots, n$), and B_0 equals his total available resources.[2] The equilibrium conditions require that

$$(2)\quad \frac{\partial P^*_1}{\partial R^*_1} \cdot S_1 = \frac{\partial P^*_2}{\partial R^*_2} \cdot S_2 = \cdots, = \frac{\partial P^*_n}{\partial R^*_n} \cdot Sn.$$

Thus, the prosecutor allocates greater resources to cases where S_i is higher, and where P^*_i is more responsive to changes in R^*_i. One would also predict charges to be dismissed against suspects where the prosecutor sees little chance of conviction regardless of his resource input, or given a conviction he expects a negligible sentence.

---

[1] See the President's Commission (1967b, pp. 4, 130–35).

[2] The price of a unit of R^*_i is assumed to be \$1.00.

Scarce resources provide an incentive for the prosecutor to negotiate a PTS with the suspect. From (1) and (2) it follows that if the prosecutor's resource inputs on a PTS equal his trial inputs, he would be willing to offer the suspect a reduced S_i in exchange for a plea of guilty, providing $P_i^* S_i$ did not decrease.[3] However, since trial costs are presumably larger than PTS costs, an even bigger reduction in S_i would be offered as this releases resources that can be used to increase the conviction probabilities in other cases. If ΔS_i denotes this sentence reduction, which is greater the greater the difference between the prosecutor's trial and PTS costs, then

$$(3) \qquad So_i = P_i^* S_i - \Delta S_i$$

where So_i is the minimum PTS sentence the prosecutor is willing to offer. In sum, the terms offered the suspect in exchange for a negotiated guilty plea will be more favorable the lower P_i^* and S_i, and the greater the prosecutor's resource saving from a PTS. Finally, note that certain cases may bring the prosecutor considerable notoriety only if a trial occurs. If this were desired, ΔS_i could also be made a negative function of notoriety, and hence for some cases ΔS_i might be negative, and So_i would be greater than $P_i^* S_i$.

The suspect's expected utility from a trial is

$$(4) \quad E(U) = P \cdot U(Wc) + (1 - P) \cdot U(Wn)$$

where P is his estimate of the probability of conviction, Wc is his endowment if convicted, and Wn is his nonconviction endowment.[4] $Wc = W - S - B - R$ where W is his endowment before any prison, bail, or other losses are deducted, and B are bail costs awaiting trial. Wc is assumed positive unless S is death, and Wn equals $Wc + S$. Since inputs of R lower P, Wc and Wn, the suspect selects a value of R that maximizes $E(U)$.[5] Given his optimal R, the suspect would reject or accept a trial depending on whether $E(U)$ were less or greater than his utility from a PTS. It is reasonable to argue that if

$$(5) \qquad \pi = U(W - So) - E(U) > 0$$

a PTS will result, since both the prosecutor and suspect can be made better off from a PTS than a trial. This follows because a negotiated sentence somewhat greater than So will leave the suspect with a utility from a PTS greater than $E(U)$, and at the same time will increase $E(C)$ for the prosecutor above its value in the event of a trial. Alternatively, if π is negative, a trial will result.

The following implications can be derived from (5) regarding the likelihood of a PTS, and the sentence resulting from a PTS.

1. Although the precise PTS sentence is indeterminant, it must lie between the extremes defined in (5). Within this range, one would expect it to depend on the relative bargaining strengths of the prosecutor and suspect. For reasons stated previously, the PTS sentence will be smaller the smaller the probability of conviction and S, and the greater the prosecutor's resource savings from a PTS.

2. π is always positive when $(W - So) > Wn$ since the PTS sentence is less than trial losses even when one is not convicted. Thus, when sentences are small relative to the costs of disputing convictions by trial, a PTS will occur.[6]

---

[3] Since sentences are set by the judge, the prosecutor cannot directly offer a lower S_i to the suspect. Instead, the suspect would be offered a conviction on a less serious charge; i.e., one that carries a lower S_i.

[4] The subscript i is deleted since it is explicit that we are now dealing with a given suspect.

[5] Maximizing $E(U)$ yields

$$(i) \quad -\frac{dP}{dR}(U(Wn) - U(Wc)) = P \cdot U'(Wc) + (1 - P)U'(Wn)$$

where U' denotes the marginal utility of the endowment in each state. The left-hand side of (i) represents marginal returns of R, and the right-hand side marginal costs of R. For a detailed analysis of the determinants of the optimal R, which analyses the effects of changes in S and W on R, see Landes (1968), pp. 23–30.

[6] This provides an obvious explanation for why many persons plead guilty to traffic violations instead of spending considerable time in traffic court disputing them.

3. If both parties agree on the probability of conviction, a PTS will take place for suspects that are risk averse or risk neutral. It can be proved that in this case a trial is equivalent to an unfair gamble.[7] Suspects maximizing their expected endowment (i.e., risk neutral) refuse this gamble, and a fortiori risk avoiders also refuse the gamble. On the other hand, risk preferers might accept this gamble. This is more likely the greater their risk preference and the smaller ΔS, B and R.[8]

4. If $P^* < P$, a trial becomes an even less favorable gamble than when $P^* = P$, and hence a PTS is more likely. The more interesting case is when $P^* > P$ because this provides an explanation in addition to risk preference of why suspects go to trial. When $P^* > P$ a trial becomes a more favorable gamble compared to $P^* = P$, and hence π falls increasing the likelihood of a trial. Moreover, it can be shown that given $P^* > P$, a trial is more likely for suspects accused of crimes that carry larger sentences.[9]

5. The smaller are B and R for a given P, the greater is $E(U)$ and the more probable a trial.[10] This suggests that policy measures designed to eliminate or reduce bail and subsidize the payment of attorney fees will have the effect of increasing the proportion of suspects who go to trial.

Court Delay. It is widely recognized that the courts are burdened with a larger volume of cases than they can efficiently handle. The results are often long delays and hasty consideration when cases reach trial.[11] This is not surprising since users pay a nominal fee, if any, and a queue develops to ration the supply.

Assume initially that there is a money price (C) for the courts, paid by the loser, that clears the market.[12] C restricts demand in two ways. First, it reduces So in (3) by a positive function of $(1-P^*)\cdot C$, which in turn raises $U(W-So)$ and π in (5) and increases the suspect's incentive to accept a PTS. Second, it lowers Wc by an amount equal to C, reducing $E(U)$, which also increases π. Moreover, since π increases as C rises, a downward sloping demand curve for the courts is generated. Further, one can venture from the analysis of (5) that as C rises the reduction in demand will be primarily from cases where there is not a significant disagreement between prosecutor and suspect over the probability of conviction and where S tends to be small. Moreover, changing the allocation of the payment of C has little effect on the above results since whether the loser or winner pays C, or both share C, a money price always increases π. Relating this to proposals to eliminate bail, it follows that no additional court demand would result if an appropriate money price were instituted at the same time bail were set at zero.

Compare this pricing scheme with one in which the courts are heavily subsidized, taking the extreme example of a zero money price. This leads to an initial increase in demand for the courts, and without an increase in supply a queue would develop. The queue imposes losses on the prosecutor as it reduces the number of convictions in any time period by delaying trial convictions. These losses increase as the queue lengthens and this in turn induces the prosecutor to offer a lower So in exchange for a guilty plea. Although $U(W-So)$ then rises, the incentive for the suspect to accept a PTS as the queue grows will depend on whether or not he is released on bail. For suspects not released, the longer the queue the higher the bail costs of a trial and hence the lower $E(U)$. This factor together with the response of the prosecutor leads to the prediction that there will be fewer trials as the queue lengthens for suspects not released on bail. On the other hand, the net effect of the queue on $E(U)$ is unclear for suspects released on bail. The loss from a given S may diminish as the penalty is pushed into the future, but being under indictment may lower the suspect's earnings. Therefore, one would predict that the demand for the courts by the latter group of suspects would be less responsive to the lengthening queue than the demand of suspects not released on bail.

---

[7] See Landes (1968), pp. 9–10.

[8] To prove this differentiate π with respect to ΔS, B and R. All partial derivatives are positive indicating that π falls with decreases in these variables.

[9] See Landes (1968), p. 11.

[10] In analyzing the effects of bail it has been assumed that B is zero for a PTS. A more appropriate assumption is that $B>0$ also for a PTS, but is on average less than bail costs of a trial. This results from the large fraction of suspects (probably more than 50 percent for suspects accused of felonies) who do not make bail, and the long queues awaiting trial. For suspects in the queue, B will be greater for a trial than a PTS. The relevant cost in the decision between a PTS and trial is the differential and for this reason we assume $B=0$ for a PTS and $B \geq 0$ for a trial.

[11] See the President's Commission (1967b, pp. 80–90).

[12] This does not mean that suspects are immediately brought to trial. Although some time is required by both suspect and prosecutor to prepare a case for trial, current delays are alleged to run considerably in excess of this. See the Commission's Report (pp. 80–90) which proposes a nine month maximum between arrest and final appeal in felony cases, in contrast to current delays that may average more than 1½ years.

Finally, it should be noted that although a zero money price is often advocated as a means of not discouraging low-income suspects from using the courts, its effect can be the opposite. A zero price operating with a bail system that tends to detain in jail low-income suspects will discourage the latter group from using the courts. In contrast, an appropriate money price will reduce the demand for the courts of suspects released on bail, and by reducing the queue can increase the use of the courts by suspects who cannot make bail. Surprisingly, in the literature that criticizes court delay, no mention is given to the possibility of charging a money price, which not only reduces delay, but can distribute more equally the courts among suspects independent of their ability to make bail.

SIMON ROTTENBERG: The allocation of law enforcement resources implies the precedent problem of defining the appropriate size of the law enforcement industry or the allocation of resources to law enforcement.

The product of that industry can be perceived to be the prevention of criminal behavior. The gain from crimes prevented is the value of harm forestalled. If this could be done at zero cost, social welfare would be maximized by preventing all crime. Since crime prevention consumes resources that have alternative uses, however, the scale of the law enforcement industry should be pushed only to the margin where harm forestalled is equal in value to the resources employed in its achievement. The empirical magnitudes of costs and gains are likely to be such that some positive quantity of crime is optimal.

Some classes of criminal activity are more harmful than others. The final unit of law enforcement resources employed in every use (in preventing every different class of crime) should forestall the same quantum of harm. This requires the calculation of harm done by each type of crime. This is not easily done. Does the cost of a given kind of crime rise at the margin? Is the utility of the pyromaniac to be netted out of the market value of the structure he burns? Are the marginal utilities of money of both victim and perpetrator to be counted in the case of coercive transfer payment as in armed robbery? How is the aesthetic disutility of nonparticipants to be costed when consensual and voluntary transactions occur in illegal commodities as in the cases of abortion, the drug trade, prostitution, and gambling? What is the social cost of usurious lending, of the employment of child labor, of the payment of less than the legal minimum wage, of tax evasion, of the contraband trade, of the sale of untaxed liquor, of the sale of goods and services by unlicensed sellers, of the sale of commodities on Sundays?

Beg the question of cost calculation and assume it is done. By distributing law enforcement resources properly, the quantity of harm done by crime can be minimized by improving the social productivity of criminal labor. Criminals are assumed to be responsive to costs imposed upon them. They will tend to choose those criminal careers in the set of all criminal careers that will maximize their expected gains net of their costs. The most important component of the cost of doing criminal business is the expected value of the punishment they suffer. That value is the product of the punishment, if inflicted, and the probability that they will be punished. The probability of punishment for engaging in any given criminal activity is a partial function of the quantity of enforcement resources devoted to identifying offenders with respect to that activity. If enforcers are more aggressive in searching out some classes of criminals and less aggressive in searching out others, they increase the cost of doing business in the former case and diminish the cost in the latter case. The allocation of the resources of the criminal justice industry among crime types, therefore, can affect the structure of criminal behavior by causing criminals to shun some kinds of criminal activity and to prefer other kinds.

The total quantity and the structural characteristics of harm done can, however, be affected not only by the size of the public crime prevention industry and the uses to which the resources it commands are put. They are also affected by the rules of the cost allocational game and by the size of the private crime prevention industry and the structure of its resource use.

Costs are imposed upon thieves, in some probability sense, when they are confronted by coalitions of police, prosecutors, judges, and wardens. But this is not the only way. They might also be imposed, for example, by triple-damage civil suits for recovery of stolen goods plus search and apprehension costs. There are, of course, many kinds of damage done which are repaired precisely in this way, and the prospect of having to bear the cost of repair induces caution and diminishes the aggregate quantity of harm done. If theft generates police activity because it has been defined as a crime rather than a civil proceeding, is it because an efficiency test has been applied and a given quantity of theft forestalled is more cheaply achieved by this strategy than by the other? Or is it because all are probabilistic victims of theft? Or because theft prevention is a public good such that a thief prevented from taking the assets of John Doe is prevented also from taking the assets of Richard Roe?

Crimes are prevented, not only by public policing, but also by private caution exercised. Safes, locks, fences, mirrors, alarms, dogs, guards, and staying at home make it costly for criminal business to be done and the more of them, the more costly it is. Expenditures of this kind by the private sector are substitutes for public crime prevention expenditures. Additional police put on patrol produce disincentives for private cautionary behavior. Given the quantity of policing done, the optimizing rules for households and firms in allocating resources for the protection of their persons and property are elementary. And the optimal public/private crime prevention magnitudinal ratio is determined on familiar cost-minimization principles.

The allocation of policing among crime types has already been mentioned and, by extension, that discussion can be applied to the allocation among neighborhoods. These do not exhaust the facets of the allocational problem.

Given a peak-and-trough time and space distribution of criminal activity, there will be periods of "down time" or waiting. It is appropriate that, in these periods, police manhours be employed in other tasks that have come to be called "public service." But if an increment of police resources is specialized to the performance of those tasks because this increases the popular propensity to come forth with information about criminal behavior, it is proper to inquire whether this information can be more cheaply bought by other techniques.

Or consider the lifetime distribution of offense. Criminal careers begin early in life. Can criminal justice allocative strategies postpone entry? If they can, whole careers of harm-doing may be forestalled. At the least, discount rates may be high enough to make the extra cost of inducing postponement worthwhile.

Or consider the alternative of preventing crime per se by patrolling and preventing it by detecting a larger proportion of all infractions and identifying and convicting a larger proportion of all offenders, thus increasing the expected value of punishment. Both techniques will diminish the quantity of crime. The rule for allocating police between these activities is clear. Its application is somewhat less clear because the empirical magnitudes of the relevant variables are not known.

It is not only in policing that allocational problems occur. They are everywhere in the criminal justice system.

For illustration, scarce prosecution and court resources can be saved from use in one subset of cases by accepting bargained guilty pleas to charges of lesser offenses and can be used in trials in another subset of cases. Scarce "correctional" resources can be saved from the care of imprisoned convicted persons and devoted to the supervision of probationers. Scarce institutional rehabilitative resources can be applied to different alternative sets of convicted persons; how should they be allocated as between old prisoners and young, first and repeated offenders, long-sentence prisoners and short, those who have assaulted persons and those who have taken another's property?

Law enforcement is an industry replete with interesting allocational problems. Economics is extraordinarily helpful in defining optimal solving rules. More so than in other industries, however, their application is difficult because many of the "commodities" that are components of cost and payoff in this industry are not market-traded. The rational administration of criminal justice requires ingenious search for proxies and the creative construction of shadow prices.

Richard B. Hoffman*: The problem of providing policing services must be considered at three levels of efficiency: the use of the existing resources, alternative allocations at the existing resource level, and the technology used to employ the resources.

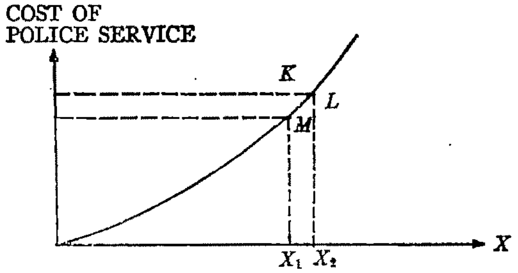


Figure 1
Effectiveness in Maintaining Public Order

Assume municipal policing agencies are attempting to maximize their effectiveness in maintaining public order; then the frontier shown in Figure 1 depicts for a given expenditure level the set of efficient operating points. However, observations on the use of policing resources indicate that policing agencies operate in the interior of feasible set; i.e., to the left of the efficient frontier.

This can be illustrated by two recent studies. The first in St. Louis indicated that if simple queuing models were used as the basis of officer assignment to surveillance and service areas in conjunction with adjusted exponentially smoothed called-for-service forecasts, improved service could be obtained on days of peak demand.[1] Resources could be shifted from other days where the work force could be reduced on a continuing basis by 15 percent to 30 percent using an assignment criteria which provided improved levels of service.[2] However, the need to meet all nonpostponeable demands for service on peak days limited the possible overall reduction to approximately 10 percent. In a second study, which the author has supervised in a major U.S. city, very similar results were found. Here, however, using the previous criteria for patrol force manning, the possible reduction in manpower appeared to be on the order of 15 to 20 percent. In this study the average demand for service in man-hours on the lowest day of the week was roughly one-third of the demand for service in on the day which had the highest demands. There were also wide variations between shifts and between districts within this city. Additionally, the number of men assigned to duty bore little relationship to the calls for service.

---

* This work was supported in part by National Aeronautics and Space Administration Grant No. NGR-05-003-125 (Technology and Urban Management) to the Space Sciences Laboratory and the Center for Planning and Development Research, Univ. of California, Berkeley.

[1] J. Thomas McEwen, "A Mathematical Model for the Prediction of Police Patrol Resources," presented to the Spring, 1968, Meeting, Operations Research Society of America, San Francisco, Calif.

[2] The assignment criterion which indicated these reductions was the immediate dispatch of a man to all major crimes and emergencies, and a probability of 95 percent that a man would be dispatched immediately to all other calls. At all times 20 percent of the patrol force would be on patrol taking no calls except in emergencies, and the remainder of the patrol force who were predominately engaged in answering calls would be patrolling at least 50 percent of the work shift.

In the face of continued demand for policing services it may be hard to understand why policing agencies have not used more sophisticated means to allocate their field manpower. When one considers that the stochastic phenomena underlying these wide variations is quite complex, it is not surprising, given the background, training, and risk-averting behavior of most policing executives, that considerable resistance could be anticipated for any changes in this direction. Further, it should be recognized that any suggestion which would prod police officers to shift their days off from society's normal Saturday-Sunday weekend, particularly those officers with long seniority, may be organizationally infeasible in a large number of departments.

From this second study we have tentatively developed some estimates of diminished patrol effectiveness caused by undermanning in precincts as a proxy for the marginal value of a patrol officer.[3] For the set of reported crimes, murder, forcible rape, robbery, aggravated assault, burglary, auto theft, and larceny over $50, the average productivity of a detective far exceeded the adjusted increase in crime[4] during periods of undermanning.[5] For example, undermanning of one man on a year-round basis (the equivalent of 4.1 man years) resulted in an estimate actual increase of 2.4 robberies while one idealized detective on a year-round forty-hour week basis solved approximately 3 robberies. The equivalent number of detectives (4.1) would solve 12.5 such crimes in a given year. Although one cannot definitely conclude that shifts from patrol to detectives would necessarily reduce criminal activity, given a model of criminal activity where total crime (C_t) is in part a function of crimes solved (C_s) and the probability of apprehension or clearance rate (C_s/C_t), experimentation in this direction would be strongly indicated.[6]

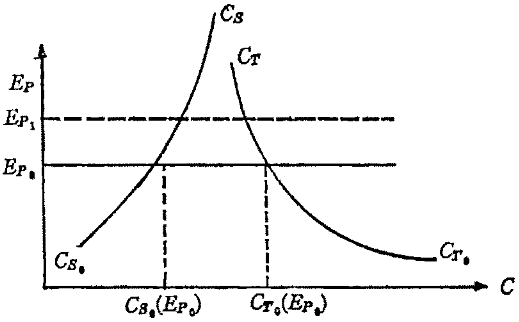


FIGURE 2

The difficulties in determining the effectiveness of police activities is further complicated by the inefficient use of existing resources. For example, it is generally agreed by both "professional" and academic authorities in the police field that administrative jobs can best be handled by civilian personnel. In many large police departments, police officers are used for report taking, answering phones, and even the typing of reports. Still in many jurisdictions, it has been estimated that the release of uniform personnel from clerical and similar administrative duties could increase the number of uniformed men undertaking police duties by 10 to 15 percent.

Reviewing the production technology, we find that the technology utilized in performing police services has undergone little if any change in the past forty to fifty years.[7] In the face of rather significant changes in the quantity, quality, and types of services demanded, the municipal police agencies have operated as if they have a homogenous labor force. The police continue to perform the same activities,

---

[3] Additionally, a recent study by the author and G. Misner cites research showing that the average police officer who is on patrol duty spends less than fifteen minutes of his working day dealing with calls relating to criminal matters. G. M. Misner and R. B. Hoffman, "Police Resource Allocation," Working Paper 73, Center for Planning and Development Research, Univ. of California, Berkeley, 1967 (presented to the 1967 Meeting, A.A.A.S., New York, N.Y.).

[4] The number of additional crimes occurring during periods of undermanning were adjusted by factors for unreported crime.

[5] The data on the average productivity of detectives from a major U.S. city is reported in detail in R. B. Hoffman, "Technology and the Provision of Policing Services," Working Paper 26 (State Univ. of New York at Buffalo, School of Management, Dec., 1968), pp. 2–3.

[6] *Ibid.*

[7] Ostensible changes have occurred, e.g., increased use of automobiles, but the use of patrol, the tasks performed while on patrol, and the methods of patrol are unchanged.

TABLE 1

MEDIAN NUMBER OF EMPLOYEES PER 1,000 POPULATION

	Population Group		
Year	Over 500,000	250,000 to 500,000	100,000 to 250,000
1940	2.23	1.35	1.47
1950	2.32	1.47	1.50
1960	2.49	1.55	1.58

SOURCE: *The Municipal Year Book*, International City Managers' Association, Washington, D.C., 1941, 1951, and 1961 editions.

detection, and patrol. Additionally, there has been no significant change in either the purposes of these activities or the "production methods" used to perform these activities.

The existing technology is described by the model in Figure 2. This model assumes that the outlay for police services is constant in the short run, and that the number of crimes which will be solved will be a function of the amount spent and the number of crimes committed. Also the actual number of crimes committed will be a function of the outlay made for police services. It may be possible to use the clearance rate as a measure of the public desire for police protection.[8] If the public felt the level of police protection was inadequate, their expenditure E_{P_0} could be increased to the level E_{P_1} reducing both the number of crimes in total and the number of uncleared crimes.

This makes crime relatively more attractive and we should not be surprised to find as noted in Table 1 that for a given city size the resources engaged in police activities have increased without any accompanying increase in effectiveness.[9]

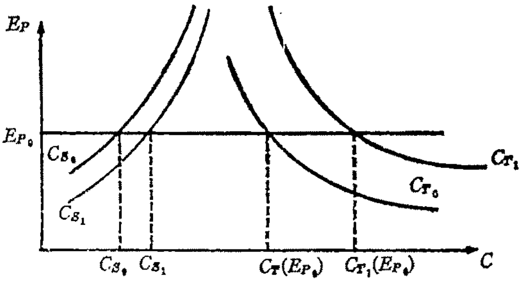


FIGURE 3

The changes in wealth and the distributions of wealth widen the opportunities for criminal activity and provide yet another force for exogenous shift in crime. Such a shift as the result of an exogenous change is shown by the model in Figure 3, with the accompanying lowered clearance rate. The unlikelihood of these forces changing direction does not provide much hope for a technology which is typified by its previous inability to control such changes.

Hence one must conclude after over thirty years of experience that the fallacy of the homogeneous labor force in police departments and the "ability of any officer to fulfill any position" imposes a very high cost on society. Without a considered reevaluation of the types of the tasks to be performed by policing agencies[10] there will be little hope of improvement. However, the truly significant improvements in productivity will come only after the proper definition and assignment of the tasks to be performed by policing agencies have been completed. If productivity is to be increased in the provision of police protection, the policing agency must begin to hire individuals who are technically competent and capable of doing the tasks which need to be performed.

[8] The fallibility of clearance data is recognized. However, for a given city the year-to-year variation would be the important factor. See J. Skolnik, *Justice Without Trial* (Wiley, 1965) for a further discussion on clearance data.

[9] This is discussed in detail in Hoffman, "Technology . . . ," *op. cit.*, pp. 9–15.

[10] Those present tasks that might be performed by other agencies and those present tasks which should not be performed at all.

CENTRALIZATION AND DECENTRALIZATION IN ECONOMIC SYSTEMS

ON THE CONCEPT AND POSSIBILITY OF INFORMATIONAL DECENTRALIZATION

By LEONID HURWICZ
University of Minnesota

The economist's interest in the issue of decentralization as against centralization is not new. It entered the eighteenth century debate concerning laissez faire versus state-imposed restrictions, as well as the nineteenth century controversies concerning the feasibility and desirability of socialism. Typically, however, the discussion was carried on with focus on concepts other than the degree of decentralization: ownership (private or collective), income determination, and the role of "value." This latter problem [11] corresponds closely to what in present-day terminology would be the issue of optimal resource allocation in a socialist economy. (The notion of optimality was implicit in the discussion rather than rigorously defined but was not incompatible with the current concepts of efficiency, Pareto optimality, etc.) It is not surprising that interest should subsequently have shifted to the optimality properties of (perfectly) competitive equilibrium which still constitutes a major aspect of currently taught economics of welfare.

Optimality Properties of Competitive Equilibria. Since the nature of equilibrium is regarded as an important feature of the economic process, the optimality properties of competitive equilibrium have been used as arguments in favor of the perfectly competitive mechanism, both in its "free enterprise" and "competitive socialist" (Lange-Lerner) interpretation. Not infrequently, the virtues of the perfectly competitive process have, by extension, been attributed (in an approximate sense) to the admittedly imperfect market economies of the Western European and U. S-type. But there continued to exist expressions of dissatisfaction both with the actual workings of the market economy and even with the perfectly competitive ideal.

The limitations of the latter have been discussed in detail in the modern literature of welfare economics. It is well known by now what assumptions concerning the economic environment underlie the basic theorems relating Pareto optimality (or efficiency) and the competitive equilibrium. To guarantee that such equilibrium will be Pareto optimal we assume absence of externalities between economic units (producers or consumers) as well as local nonsaturation [8]. (These are the "direct" or, on production side, "technological" externalities, as distinct from what Viner and Scitovsky call "pecuniary" externalities [14].) To guarantee the possibility of attaining every Pareto optimal allocation through the competitive mechanism, more assumptions are required, among them absence of indivisibilities and convexity both on consumption and on production sides, the latter ruling out the presence of increasing returns [8].

Nonclassical Environments. If one is unwilling to postulate that the real economic environment satisfies these "classical" assumptions, one is naturally led to the consideration of alternatives or modifications of the perfectly competitive mechanism that would perform satisfactorily not only within the classical environments but also in the presence of such nonclassical environmental features as increasing returns, indivisibilities, etc. Both Marshall's proposal for the subsidization of increasing returns industries and the marginal cost pricing ideas of the 1930's (Hotelling, Lange, Lerner) are examples of attempts to modify the competitive mechanism to cope with nonclassical environments. (An alternative, imperfectly competitive, mechanism to bring about optimality in the presence of increasing returns was proposed by Arrow and Hurwicz [1].) Similarly, there are in the economic literature suggestions for dealing with situations involving various combinations of indivisibility, externality, and public goods, by calculation techniques based on, or related to, consumers' (and producers') surplus (Dupuit, Hotelling, Lerner).

Formula versus Mechanism. But the existence of a mathematical formula characterizing (usually only implicitly) optimal solutions does not imply the existence of a mechanism for processing the relevant information to the point of obtaining even reasonable approximations to such solutions. This was the point stressed by Robbins [12, p. 151] and Hayek [4, p. 209–12]. The issues raised by them, especially those by Hayek (subsequently elaborated in 1945 and 1946), are of crucial importance in the present context, since they have to do with the centralized nature of socialism (collectivism, planning) as against the decentralized nature of the competitive market process. By analyzing these issues, we may arrive at the notions of decentralization and centralization implicit in the debate. In this writer's view these notions point to important aspects of the problem and can be built upon in considering the problem of feasibility of decentralization in the context of recent controversies.

Dispersion of Information. The fundamental point made by Hayek with regard to availability of economic information is its dispersion. A simple way of visualizing this dispersion is to assume that, at least at the beginning of the allocative process, each economic unit has perfect information concerning itself (its production function or its preference map or its resource holdings) and no information concerning the other units. During the allocative process there will be communication, then decisions, and finally their implementation. An oversimplified view of a centralized process envisions the transfer of the dispersed information to a single unit (say the central planning board) which then performs calculations yielding appropriate actions (especially in the production and input flows) to be taken by the other units. It is this transfer of dispersed information to the central unit that Hayek considered impossible. (The magnitude of calculations that would have to be performed centrally was another difficulty.) And it was the need for such a transfer that was for him, implicitly, the characteristic feature of a centralized allocative process.

Transfer of Information. Now things would be easy if one could define a decentralized allocative process as one in which no information is transferred from any unit to any other unit. We would then have two polar extremes: (informational) centralization, with complete transfer of all dispersed information to a single designated unit, and (informational) decentralization, with no transfer of information as between units. Such a definition of decentralization could be adopted if there were no communication among the economic units during the allocative process. But, unless we want to disqualify the competitive process from the decentralized class, we cannot prohibit all communication. Hayek who was perhaps the first to describe the competitive process as decentralized ([4, p. 211] [6, p. 524]; also [5, pp. 48–50]) stresses the need for communication among units to achieve co-ordination of their actions [6, pp. 525–27]. This communication is accomplished in the competitive mechanism through bids (representing quantities supplied or demanded) and prices attached to those bids.

Communication in the Competitive Process. There is an intuitively appealing difference between the type of communication that prevails in a competitive market (where bids and prices are communicated) and the type of communication envisaged in the central planning board model (where complete descriptions of technologies or preferences or resource holdings are communicated). This difference can be used as a basis for classifying allocative mechanisms as centralized, decentralized, or in between, depending on whether their communication requirements are like those of the competitive mechanism or like those of the above central planning board model or in between. But what type of communication would we accept as being "like that of the competitive mechanism"? If our definition is too narrow, we shall be going in the direction of almost identifying the notion of informational decentralization with that of a "price mechanism." Yet [10, p. 204] the intuitive concept of informational decentralization seems broader than that of a price mechanism (as yet not rigorously defined either). On the other hand, if virtually any type of communication were considered acceptable, we might find that even the central planning board model would qualify as decentralized, thus failing to express the difference between the two types of communication.

To arrive at a possible answer let us see first what kind of communication takes place in the competitive process. Geometrically, a bid is a point in the commodity space; algebraically, it is a vector with components corresponding to the goods present in the economy. (A typical bid would have most components zero, with the rest positive and negative.) A price can be represented either as a half-space in the commodity space or as a point, algebraically again a vector of the dimensionality of the commodity space. (This point or vector representation of prices is not generally valid for infinite-dimensional commodity spaces.)

A Concept of Informational Decentralization. This suggests that we accept as "like competitive" those processes in which messages consist of vectors whose dimensionality is the same as that of the commodity space (assuming the latter to be finite). Suppose that we make this requirement a part of our definition of (informational) decentralization. Our minimum desiderata will be satisfied if the competitive process does qualify while the central planning board process does not. Now clearly, the competitive process does qualify. But have we disqualified the central planning board process?

To answer this question we must find out whether a message consisting of as many (real) numbers as there are goods can be used to transfer such information as the complete description of, say, a production function. Perhaps somewhat unexpectedly, under certain circumstances the answer may be affirmative. To begin with, and this is least surprising, the information could easily be transferred if, for instance, all production functions were in advance known to belong to a parametric family (e.g., Cobb-Douglas) with no more parameters than there are goods, since in this case the components of the transmitted vector message could be reserved (in a prearranged manner) for these parameters. But in fact it does not matter if there are more parameters to be transmitted than there are goods in the economy! The reason is the possibility of using one real number to transmit information about several (arbitrarily, but finitely, many) numbers. (Suppose that one decimal number, say $0.x_1x_2x_3x_4 \cdots$ is to carry information about two decimal numbers, say $0.y_1y_2 \cdots$ and $0.z_1z_2 \cdots$; then we merely reserve the odd-place digits of x for the y digits and the even-place digits of x for the z digits: $x_1=y_1$, $x_2=z_1$, $x_3=y_2$, $x_4=z_2$, etc.)

A Priori Admissible Environments. Hence in an economic environment where production functions and utility functions were representable by finite-parameter families of functions, the seemingly reasonable communicational characterization of decentralization (viz., that only commodity-dimensional messages can be transmitted) turns out inadequate because it would not screen out the par excellence centralized process! Should we worry about this? Possibly not. For we may be looking for a process that would guarantee optimal resource allocation even when the variability in the economic environment exceeds any finite-parameter families of functions. For instance, suppose that a firm's production possibility set could be any subset whatsoever of the commodity space. In that case the collection of all commodity-dimensional vectors is known (from set theory) to be inadequate as a representation of all such production possibility sets. (The same would hold true if we confined ourselves to a somewhat narrower class of production possibility sets satisfying certain conventional assumptions.)

Thus when the a priori admissible class of production sets (or preference maps) is sufficiently broad to be beyond finite-parameter description, the central planning board process no longer qualifies as decentralized if communication in the latter is confined to commodity-dimensional messages. Since the competitive process still does qualify, our minimum desiderata for a definition of informational decentralization are satisfied.

Let us then adopt a definition of informational decentralization which restricts communication to commodity-dimensional messages and also postulate that the only information available to any economic unit concerning the other units is derived from such communication; i.e., except for what can be inferred from such communication, every unit is assumed to be in total ignorance of other units' technologies, preferences, and resource holdings. Stated in this manner, the restrictions may appear overly severe, and certainly unrealistic as descriptions of phenomena observed in the real market place. But these restrictions are satisfied by the idealized Walrasian model of the competitive mechanism and in this sense are not too severe.

Of course, this definition would be of little interest if the competitive mechanism were the only one to qualify under it. But in fact, a number of other proposed mechanisms also qualify, especially when above communication restrictions are interpreted as permitting directed communication; i.e., communication where in each phase of the allocative process some units are permitted to send different messages to the various other units, each message still being required to be a commodity-dimensional vector. (Since the competitive process can get along without directed messages, it becomes clear that the concept of informational decentralization we are now using calls for less informational efficiency than characterizes the competitive process. A more stringent definition was used by the writer in 1959.) In particular, "decentralized planning" procedures proposed by Malinvaud [9, Sections III, IV, V] as well as the Kornai-Liptak process satisfy (with certain qualifications to be stated below) the requirements just formulated. (In the Kornai-Liptak process there is a central body which receives from the various production units their "local" shadow price messages and responds with proposed input-output vectors. Such a process does not satisfy the more stringent communicational requirements of self-relevance and anonymity in this writer's 1959 definition.)

Performance Criteria. We shall not at this point pursue the different possible variants of the decentralization concept and their respective merits. It is more interesting to see what questions can be asked given *a* (not *the*) concept of informational decentralization. In what sense is it meaningful to ask whether such decentralization is "possible"? We have seen already that many processes satisfying the requirements of our definition can be defined. (This remains true even for more stringent variants of the concept.) Thus our use of the term possible in this context is somewhat misleading. What we want to know is not merely whether such a process can be defined, but whether it will have certain optimality properties with regard to the allocations it produces. What properties should we specify?

Pareto Satisfactoriness. Here again, we get some guidance from the situation prevailing in the "classical" case. The economists have often been satisfied if they could prove that every equilibrium is (Pareto) optimal and every optimum capable of sustaining an equilibrium. The conjunction of these two properties (together with a sort of single valuedness which is also present in competitive equilibria) has been labeled "Pareto satisfactoriness" of a process. Thus if we want an allocative process to possess some of the important static attributes of perfect competition operating in a classical environment, we may require that it be Pareto satisfactory. With this interpretation of what we mean by the possibility of decentralization, it is evident that this possibility may depend on the (a priori admissible) class of environments the mechanism is expected to operate in. For instance, the competitive mechanism qualifies as informationally decentralized and is known to be Pareto satisfactory within the family of classical environments (no externalities, no indivisibilities, convexity). But what if our interest goes beyond these classical environments?

For instance, suppose cases of increasing returns are members of the a priori admissible class. We know that the competitive mechanism, although still decentralized, is no longer Pareto satisfactory. What about Marshall's subsidy scheme? This may conceivably be Pareto satisfactory, but appears to violate the communicational requirements of decentralization, since the subsidy level decisions might require detailed knowledge of the production functions on the part of the subsidy granting body. (The preceding statement is merely a conjecture. A rigorous study remains to be carried out.)

It is seen that the possibility of a decentralized mechanism outside the classical environment sphere is by no means obvious. To expect that there would exist mechanisms both decentralized and Pareto satisfactory for all conceivable environments does not seem reasonable. But what about the class of all environments that are free of externalities? This class contains some of the most troublesome cases (increasing returns, indivisibilities, discontinuities of preferences). Hence it is perhaps somewhat surprising that decentralization within this broad class of environments is possible. (The latter statement requires qualification. Two processes have been designed to cope with the class of all environments that are free of externalities. The earlier of the two is presented under the name of "greed process" in the writer's 1959 paper. This process is Pareto satisfactory in the sense of the preceding definition but qualifies as informationally decentralized if a message is permitted to contain not merely a single commodity-dimensional vector but also a set of such vectors; i.e., in this respect the 1959 definition is less stringent than the present one, although it is more stringent in other respects: anonymity and self-relevance. The more recent process, to be published in a paper by Hurwicz, Radner, and Reiter, satisfies the more stringent informational requirements—in particular, messages are single vectors, and anonymity and self-relevance hold—but it excludes discontinuous divisible environments, while permitting indivisibilities and nonconvexities, and is probabilistic in nature. Hence so is its Pareto satisfactoriness. On the other hand, it has, unlike the greed process, good convergence properties as well.)

Malinvaud's Criteria. If one thinks of actually using a given allocative process to arrive at a decision, one is no longer satisfied by Pareto satisfactoriness, even if accompanied by convergence (stability of equilibrium). In fact, one may be willing to sacrifice the good static features of equilibria and convergence (as the number of iterations tends to infinity) in exchange for some needed features obtainable after a finite number of iterations. Malinvaud [9] has focused his attention on requirements of this type. Thus, for instance, he requires that the solution obtained after a finite number of iterations be feasible in terms of the available resources and technology. Malinvaud points out that certain Walrasian tatonnement procedures related to gradient processes need not yield feasible plans if interrupted after a finite number of iterations. (These are procedures based on the work of Arrow, Hurwicz, and Uzawa.)

Thus, it is not obvious that informational decentralization is possible (in the sense of Malinvaud's performance requirements) even for all classical environments. Malinvaud's paper shows that two processes satisfy his performance requirements for certain narrower classes of environments. The Taylor process [9, Section IV] satisfies Malinvaud's performance requirements for the class of the Leontief-Samuelson environments. His other process [9, Section V] satisfies his performance requirements when certain conditions of convexity, closedness and boundedness are satisfied by the environment. (The class of environments covered is not much narrower than the classical. However, the informational requirements of this process are much more stringent than those of the competitive process in that cumulative "memory" is postulated for the planning bureau.) Unfortunately, Malinvaud is "buying" the performance properties he is interested in (in particular, the feasibility of the final plan) by a weakening of the informational decentralization aspect; viz., by his assumption that the planning bureau has initial knowledge of a feasible plan. This assumption violates one aspect of our definition of informational decentralization. If one were to abandon Malinvaud's assumption that the planning bureau has such initial knowledge, his processes [9, Section IV, V] will qualify as informationally decentralized under our definition but may no longer satisfy Malinvaud's performance criteria.

If we insist on our definition of decentralization, it is still not obvious whether decentralization is possible in terms of Malinvaud's performance criteria even for, say, a class as narrow as that of Leontief-Samuelson environments. (We have not as yet examined the Kornai-Liptak process from this point of view.)

Decentralizability of Nonpecuniary Externalities. Let us now return to (static) Pareto satisfactoriness as our performance criterion. Subject to this criterion, what are the limits of informational decentralizability; i.e., for how broad a class of environments could there be an allocative mechanism that is both informationally decentralized and Pareto satisfactory? We have seen already that the class of all or most environments that are free of externalities can be taken care of either by the (set-valued) greed process or its stochastic (point-valued) counterpart. But what about environments that do have externalities?

In the last decade there have been several important contributions that have bearing on this issue. The most interesting from our point of view is the exchange between Wellisz [17] and Davis and Whinston [3] [4]. A central issue in the debate was the feasibility of the Pigou system of taxes and subsidies to supplement the workings of the market so as to produce optimality. Davis and Whinston, in their earlier paper [2], pointed out some of the difficulties of treating the problem of externalities in the Pigovian fashion, specifically when these externalities are of the bilateral nonseparable type. Wellisz, in addition to questioning some of their arguments, proposed what he called a "Pigovian solution to the problem of externalities." Denoting by $C_i(q_i, q_j)$, $j \neq i$, the total cost function of the ith firm, $i=1, 2$, with the presence of q_j indicating the externality, the Wellisz approach consists in finding two functional relations g_i, $i=1, 2$, with whose help he constructs the modified ("internalized") cost functions $C_i'(q_i)=C_i[q_i, g_i(q_i)]$. (Although Wellisz refers to g_i as the "Pigovian taxes," a better interpretation would seem to be that of "conjectured response functions." I.e., $g_i(q_i)$ would yield the value of q_j conjectured by unit i to be j's response to i's choosing quantity level q_i. The difficulty with the tax interpretation can be seen in the separable case where, say, $C_1(q_1, q_2) = C_{11}(q_1)+C_{12}(q_2)$. The modified cost function is here given by $C_1'=C_{11}(q_1) + C_{12}[g_1(q_1)]$. Hence the amount of tax paid would seem to be $C_{12}[g_1(q_1)]$ rather than $g_1(q_1)$. It is also worth noting that g_1 and g_2 are mutually inverse.)

The basic Davis-Whinston objection to the Wellisz solution is that the calculation of each of the two "response functions" g_i requires the knowledge of both cost functions C_1 and C_2. Thus if the ith firm were to compute its g_i, it would also have to know the other firm's cost function C_j, $j\neq i$. If one assumes that initially information is "dispersed," so that each firm only knows its own cost function, the calculation required by the Wellisz proposal could only be carried out if the relevant information about the other firm's were transferred through appropriate communication; clearly, this does not seem to be in the spirit of informational decentralization as defined in the present paper. Davis and Whinston, in fact, claim that "the informational requirements of the Wellisz proposal are just as stringent as those of a centralized scheme" [3, p. 310] or "of a (partially) centralized scheme" (p. 312). However, their attempt to establish this claim is vitiated by the absence of a rigorous definition of what type of communication they regard as corresponding either to (informationally) centralized or "partially centralized" or decentralized schemes. (A further weakness of their exposition is due to the fact that an example designed to show the difficulties of the Wellisz proposal does not possess a joint optimum. In fact, the joint profit function given in their equation (19), p. 311, as $P=3q_1+4q_2 -(q_1+q_2)-(q_1q_2)$ can be made arbitrarily large by setting $q_1=0$ letting q_2 tend to infinity. However, the phenomenon which the example is meant to illustrate could probably be exhibited in the context of another similar example possessing a finite optimum.) According to Davis and Whinston, it is "in general" necessary to find the optimizing values of the respective outputs q_1 and q_2 in order to calculate the "response functions" g_i. If this is so, they say, one might as well have a central authority dictating these outputs to the two firms rather than helping them "internalize" their cost functions. However, their example (aside from the above weakness) only shows that the calculation of optimal outputs is sufficient to obtain the response functions; the necessity is not rigorously established. Since there is a variety of possible "adjustment processes" (Hurwicz, 1959) corresponding to the Wellisz scheme, one would, presumably, have to show that none of those can satisfy the requirements of a specific definition of informational decentralization, e.g., of that given in the present paper, or some other which may have been implicit in the Davis and Whinston arguments. This has not as yet been done.

Whinston's Scheme. Having adopted the (admittedly plausible) conclusion that the Wellisz scheme does not qualify as informationally decentralized, Davis and Whinston propose a very interesting alternative (originally due to Whinston [18] which in their view does qualify. Their scheme is valuable on its own merits, but is of particular interest here as throwing light on what one may mean by informational decentralization. The basic idea of the proposal is to give each firm two decision variables: the ith firm chooses not only its own output q_i but also the output x_j it proposes for the other firm. A third party ("the authority") sets unit charges to be paid by one firm to the other (in the sense of tatonnement, with payments only at the end of iteration or in equilibrium). These charges are proportionate to the discrepancy between q_i and x_i, with the level of charges set by the authority according to a formula also involving the size of the discrepancy. This process appears to have satisfactory static properties, although its stability properties are not clear to this writer. But our main interest here is in its informational features. The authors state that "it does not require that information concerning cost functions be communicated to any party (including the government). In this respect it is informationally decentralized. All that is required (aside from the requirements that firms know their own cost functions, etc.) is that the existence of externalities be recognized" [3, p. 312]. "Similarly, the governmental authority does not need to have a knowledge of the cost curves of the two firms. It simply adjusts the per unit charges according to the stated rules" [3, pp. 315–16].

The notion of informational decentralization implicit in these statements is in accord with that of our earlier discussion: in a decentralized allocative process there should not be transmission of detailed information concerning one unit (cost functions, production functions) from that unit to another unit whether the latter be a firm or some third party (in this case a government authority). Since it is recognized, however, that some communication must take place between units to ensure optimality (or even feasibility) and that such communication is bound, indirectly at least, to convey some information about the unit, we are again faced with the problem of what kind of communication will qualify as still informationally decentralized.

The characteristic feature of the Davis-Whinston proposal is that each firm has two decision variables (its own output and the output it proposes for the other firm) rather than merely one. In addition, the third party (the government authority) is endowed with new variables under its control; viz., the two discrepancy charges. One is entitled to ask whether there is no limit on the proliferation of the number of variables before we conclude that the mechanism no longer qualifies as informationally decentralized.

The terminological label is not the important thing, however. It may seem artificial to decide precisely where the boundary is drawn. The significant aspect of the example is that the (Davis-Whinston) scheme proposed does have higher communicational requirements than the corresponding environment would have had if the external effects had been eliminated. From this point of view the interesting question to ask is whether this increase in communicational requirements is unavoidable, and if so, to what extent.

That the presence of externalities would call for more communication is intuitively clear, since the same action by a given unit may or may not be optimal (or even feasible) depending on the other unit's behavior. But care must be exercised in formalizing the nature of this concept of "more information." Mere counting of the number of transmitted variables (or messages) is inadequate for reasons indicated earlier in this paper; viz., that it is possible to have one number serve as a vehicle for transmitting several. It would appear that the proper way to rule out such seeming paradoxes is to impose restrictions on the continuity (or smoothness) of the functions representing the responses (decision making) of the participants in the allocative process. (Results along these lines were presented by this writer in his 1966 paper on the decentralizability of externalities.) While such mathematical restrictions may at first lack intuitive appeal, they result, in the writer's opinion, in a closer approximation to the intuitive notion of the dimensionality of the message than mere counting of numbers contained in it. Along these lines, the two questions would be whether the Davis-Whinston scheme calls for more information transfer than would have been required in the absence of externalities and, assuming the answer to be in the affirmative, whether one could not get by with less.

Commodity Interpretation. As to the latter question, it should first be noted that the additionally introduced variables of the Davis-Whinston scheme are quite natural if one views the phenomenon of externality as the presence of an additional commodity which is being ignored by the usual market mechanism. For instance, if the externality is due to smoke being blown from factory 1 to factory 2, we may think of this (in the absence of zoning prohibitions) as a commodity (smoke) being "sold" by factory 1 to factory 2 without the latter's consent. If smoke were declared a commodity like any other, factory 1 could not be forced to "buy" smoke without its consent; presumably it might do so at some negative price. Thus by declaring smoke to be a commodity we would be introducing two more variables into the system; viz., the price of smoke and the quantity of smoke demanded by the second factory. This agrees with the two extra variables per externality in the Davis-Whinston scheme.

The smoke example points, however, to a possible improvement of the Davis-Whinston scheme in the sense that the variable "quantity of smoke demanded by 2" has lesser informational requirements than the Davis-Whinston variable "level of output in 1 proposed by 2." For when 2 talks only about the amount of smoke it will receive, it is up to factory 1 to determine the relationship between its level of output and the amount of smoke emitted. (The formal model underlying this remark is one in which the cost function of the second factory is, say, $F_2(s_1, q_2)$ while $s_1=\phi_1(q_1)$ where s_1 represents the amount of smoke sent from 1 to 2 and we assume that factory 2 only knows F_2 while factory 1 knows the technological relation ϕ_1.)

An Alternative Scheme. While it is thus seen that the Davis-Whinston scheme is consistent with a traditional view of externalities as involving commodities that bypass the market process, it is still legitimate to ask whether a mechanism could not be devised involving fewer variables. To explore this in the spirit of the example considered by Davis and Whinston and in much of the other literature, and without claiming generality, we consider two firms with total net profit functions $f^i(q_1, q_2)$, $i=1, 2$, the functions being assumed differentiable and concave so that joint profit (i.e., the maximum of f^1+f^2) is obtained by setting the sum of appropriate partial derivatives equal to zero. I.e., the optimality conditions are

$$f^1_1(q_1, q_2) + f^2_1(q_1, q_2) = 0$$

$$f^1_2(q_1, q_2) + f^2_2(q_1, q_2) = 0$$

where the subscript j denotes the partial differentiation with respect to q_j.

We shall now consider a communication scheme in which the ith firm communicates f^i_j, for $i \neq j$, to the other firm through a message denoted by m^i_j. (This message is a single number.) Its own proposed level of output is denoted by m^i_i. With t as the variable of iteration ("time") we consider the following adjustment process:

Messages sent out by firm 1:

(1.1) the equation

$$f^1_1[m^1_1(t), m^2_1(t-1)] + m^2_1(t-1) = 0$$

is solved by firm 1 to determine $m^1_1(t)$;

(1.2) the equation

$$m^1_2(t) = f^1_2[m^1_1(t), m^2_2(t-1)]$$

is used by firm 1 to determine $m^1_2(t)$.

Similarly, for firm 2:

(2.1) the message $m^2_2(t)$ is obtained by solving equation

$$m^1_2(t-1) + f^2_2[m^1_1(t-1), m^2_2(t)] = 0;$$

(2.2) the equation

$$m^2_1(t) = f^2_1[m^1_1(t-1), m^2_2(t)]$$

is used to obtain $m^2_1(t)$.

(Note: in the right members of (1.2) and (2.1), $m^i_i(t)$ could be replaced by $m^i_i(t-1)$.)

A modification along the lines of the smoke example above could be carried out as follows. Assume that, say, $f^2(q_1, q_2)$ can be expressed as $F^2(s_1, q_2)$ where $s_1=\phi^1(q_1)$, with similar relations in the other direction.

Then the relations (2.1)-(2.2) would respectively be replaced by the following

$$(2.1') \quad m^1_2(t-1)\phi^2_2[m^2_2(t-1)] + F^2_2[s_1(t-1), m^2_2(t)] = 0$$

$$(2.2') \quad m^2_1(t) = F^2_1[s_1(t-1), m^2_2(t)].$$

Equations (1-1)-(1.2)′ would be replaced by analogous (1.1)′-(1-2)′.

Not surprisingly, it is seen that the terms involving m^i_j, $i \neq j$, reflect the marginal externalities (marginal external costs or marginal productivity of the "input" smoke). An interpretation in terms of "demand price" for smoke, etc., seems possible.

It should be noted that neither system involves an outside authority. Furthermore, while the "smoke model" (with primes) involves six variables (the m^i_j's and the s_i's), only four of these are involved in outside communication (the s_i's and the m^i_j's with $i \neq j$). Hence the communication requirements would in both schemes be lower than in the Davis-Whinston proposal. It should be made clear, however, that this does not necessarily imply an overall superiority of the former over the latter because of possible increase in computational complexity. The purpose of our construction is merely to push somewhat further the exploration of the possibilities of saving on the extent of communication between the economic units in the presence of externalities and also to see to what extent one can get by without introducing third parties. (But let us remember that this paper ignores issues of motivation and enforcement which might require the intervention of third parties for reasons other than communication.)

The alternative schemes just proposed still fail to answer the question of how much "better" one could do in economizing on the number of messages to be transmitted. Always bearing in mind that the message number counting must be done subject to "smoothness assumptions," a partial answer is provided in the writer's 1966 paper where it is shown that even for quadratic profit functions two messages would, in general, be insufficient when bilateral externalities are present. (The smoothness requirements can be taken to mean the Lipschitzian property of the decision rules; mere continuity is not enough.)

A reference should be made to the existence of other proposals in the direction of informational decentralization in situations involving externalities. In particular, an ingenious extension of the greed process is contained in an unpublished paper by A. Camacho (1967). Its communicational requirements, however, go beyond what would qualify as informationally decentralized both under the standards of the present paper (because set-valued messages are used, as in the greed process) and under those of the 1959 paper (because self-relevance of messages appears to be violated). It is nevertheless conceivable that the Camacho scheme ("D-process") is close to being maximally informationally efficient for the class of environments it deals with. Additional research is needed in this direction.

Pecuniary Externalities. A discussion of the issue of decentralization in the presence of externalities would be incomplete without at least a brief consideration of the problem of the "pecuniary" externalitites. In fact, a perusal of the existing literature [13] [14] [15] shows that the pecuniary externalities are a major reason given for considering planning as superior to the market mechanism in certain situations, especially in underdeveloped economies. A detailed analysis of the problem would exceed the scope of this paper, and we shall confine ourselves to a few comments.

Since a great many different phenomena are lumped under the pecuniary externalities label, let us limit ourselves to the case (Allyn Young, Rosenstein-Rodan) of the problem where two capital goods (a subway-"tube" and a housing development) whose services (transportation, shelter) are complementary. The claim made in favor of "planning" (or merger)—i.e., against decentralization of the investment decisions—is that there is uncertainty on each side as to the decision of the other. Since our focus is on the possibility of informational decentralization, it is natural to ask first whether the competitive model (as distinct from the actual market structure) could be used here. The answer is in the affirmative if the complementarity of the services is the only "troublesome" aspect of the situation. What is then needed, however, is the existence of a market for such services, so that at the time investment (construction) decisions are made there exist prices equilibrating the market for such services, to be taken into account in the investment decisions. But the services are in the future. Hence a market for future goods would have to be created, possibly in a "simulated" or "shadow" form. (A suggestion to this effect by K. J. Arrow is quoted in [4, p. 306].) From the informational point of view this would be acceptable as a decentralized solution. (The objections [16] [15] to the centralized planning solution for such situations point out the possibility that economic units might have information concerning the units which might undertake construction resulting in satisfying the requirements of complementarity, thus reducing uncertainty. This is not to be denied as an empirical fact, but it should be noted that the availability of such information violates the postulate of dispersion of information and thus that of informational decentralization. As to models involving expected prices of future goods, the difficulty is that (*a*) not all market participants would necessarily have the same expectations and (*b*) even if their expectations were unanimous, the outputs decided on the basis of such expectations might fail to equate future demand and supply. In any case, one should keep in mind the fact that optimality here must be defined as *ex ante*, not *ex post*.)

The preceding paragraph was addressed to the case where complementarity of services was the only "trouble." In fact, investment decisions may also involve indivisibilities or increasing returns. Here we are back to the difficulties of nonclassical environments discussed earlier.

REFERENCES

1. KENNETH J. ARROW AND LEONID HURWICZ, "Decentralization and Computation in Resource Allocation," reprinted from *Essays in Economics and Econometrics* (Univ. of North Carolina Press, 1960), pp. 34–104.
2. OTTO A. DAVIS AND ANDREW WHINSTON, "Externalities, Welfare, and the Theory of Games," *J. P. E.*, June, 1962, pp. 241–62.
3. ———, "On Externalities, Information and the Government-Assisted Invisible Hand," *Economica*, Aug., 1966, pp. 303–18.
4. F. A. HAYEK, ed., *Collectivist Economic Planning* (London: Routledge and Kegan Paul Ltd., 1935).
5. ———, *The Road to Serfdom* (London, 1946).
6. ———, "The Use of Knowledge in Society," *A.E.R.*, 1945, pp. 519–30.
7. L. HURWICZ, "Conditions for Economic Efficiency of Centralized and Decentralized Structures," in G. Grossman, ed., *Value and Plan* (Univ. of California Press, 1960), pp. 162–83.
8. T. C. KOOPMANS, *Three Essays on the State of Economic Science* (McGraw-Hill, 1957).
9. E. MALINVAUD, "Decentralized Procedures for Planning," in E. Malinvaud and M. O. L. Bacharach, eds., *Activity Analysis in the Theory of Growth and Planning* (London: Macmillan, 1967), pp. 170–208.
10. JAMES MARCH AND HERBERT SIMON, *Organizations* (Wiley, 1958).
11. N. G. PIERSON, "The Problem of Value in the Socialist Society," reprinted in *Collectivist Economic Planning*, F. A. von Hayek, ed. (London: Routledge, 1935).
12. LIONEL C. ROBBINS, *The Great Depression* (New York: Macmillan & Co., 1934).
13. P. N. ROSENSTEIN-RODAN, "Problems of Industrialization of Eastern and South-Eastern Europe," *Econ. J.*, 1943, pp. 202–11.
14. TIBOR SCITOVSKY, "Two Concepts of External Economies," *J.P.E.*, Apr., 1954, pp. 143–51.
15. GERALD SIRKIN, *The Visible Hand: The Fundamentals of Economic Planning* (McGraw-Hill, 1968).
16. J. A. STOCKFISCH, "External Economies, Investment, and Foresight," *J.P.E.*, Oct., 1955, pp. 446–49.
17. STANISLAW WELLISZ, "On External Diseconomies and the Government-Assisted Invisible Hand," *Economica*, 1964, pp. 345–62.
18. A. B. WHINSTON, "Price Guides in Decentralized Organization," in William W. Cooper, Harold J. Leavitt, and Maynard W. Shelley, II, eds., *New Perspectives in Organization Research* (New York, 1964).

ON THE COMPARISON OF CENTRALIZED AND DECENTRALIZED ECONOMIES

By THOMAS MARSCHAK
University of California, Berkeley

To understand when and why one mechanism for the allocation of resources in an economy performs better than another appears to be the slippery and distant goal of two groups of patient investigators. In pursuing the goal both groups slowly hack their way through a common jungle but the tools the two groups use and the language they speak are typically so different that they seldom meet to share their experiences and to learn from each other. One group studies existing economic systems; it looks at the complex noneconomic determinants of economic performance but also tries to arrange the purely economic data, often very scarce, into patterns that are suggestive for comparisons of system performance. The other group formulates abstract models of economic systems, seeks to make precise the meaning of "better" performance and tries, at least for simple examples, to find interesting conditions on two systems which imply better performance of one of them. As a wavering member of the second group I shall try in these brief remarks to suggest some ways in which empirical comparisons—those feasible now and those which would be feasible with more data—might to some extent be guided and clarified by the study of abstract systems.

Centralization and Decentralization. The terms "centralized" and "decentralized" pervade the comparative study of economic systems. Perhaps if these terms could be erased from the language much misunderstanding and oversimplification would disappear. Doubtless, however, we are stuck with these terms, and they are commonly used, as we shall see, in stating the major general questions facing comparative economics. Their wide use might possibly be turned into something helpful if careful definitions were widely accepted as well. At the risk of impinging on Professor Hurwicz' paper, I shall quickly suggest one.

What are the objects, first of all, that may have the property of being centralized or decentralized? The terms "mechanism," "system," or "adjustment process" have been used for these objects. For the moment I propose to use the simple term "process." A process is used by an economy and proceeds in a sequence of steps. An economy consists of a number of agents. Each agent has at the start of any step a vector of states. Each state may be thought of as a "slot" containing information which changes from step to step. During a step an agent receives signals from other agents and from the economy's environment. The environment includes the economy's primary resource endowments, technology, and tastes. To receive a signal from the environment is to perform an observation on it. Given any environment, the signal received from it by each agent (his "natural" observation on the environment) is the same whatever process is used. Given his states at the start of a step and the signals received during the step, each agent sends signals to other agents at the end of the step and also modifies his states.

Thus each agent's behavior is completely defined by two functions: one function is from received signals and current states to the new states; the other is from received signals and current states to the new outgoing signals. A process is a collection of pairs of these functions, one pair for each agent.[1] All but one of an agent's states merely serve to store information. The remaining state, called an "action," expresses the agent's physical acts currently in force (e.g., the productions and trades in which the agent is currently engaged). The time required for an agent to complete a step—to perform the computation, communication, observation, and storage which it requires—is to some extent subject to choice: more elaborate information processing equipment can shorten the time.

In the simplest case the sequence of environments confronting the economy is unaffected by the behavior of agents and follows a known probability law. The net performance of a process over time is then a function of the sequence of environments, the sequence of actions, the step completion times, and the resources required to operate the process while achieving these step completion times. Specifically we may take this function to be the discounted value of the payoff which accumulates over time as the process is operated. Payoff is defined for each instant of time; it is a function of the actions in force and the environment prevailing at that instant. The gross performance of a process (useful for some purposes) is the value of net performance when it is assumed that no resources are required to operate the process while achieving the chosen completion times. To correctly rank two processes, a heroic calculation is required: that process is preferred for which expected net performance is larger when for each process one chooses step completion times that maximize expected net performance. Quicker step completion generally improves gross performance, since it means that actions are a less delayed response to the changing environment so that higher payoffs are obtained; whether it also improves net performance depends on what additional resources are required to shorten the step completion time. Short of the heroic calculation, one may still be able to make good inferences about expected net performance of a process by observing a short sequence of environments and the accompanying sequence of payoffs which the process generates. Possible payoffs include per capita national product and per capita consumption.

The concept of process just given is broad enough to include idealizations of competition, of central planning (one of the agents being a central planning office engaged in interchange with managers and finally determining a plan which specifies the other agents' actions), of a Lange-Lerner scheme with central price-setting, and many other mechanisms. It is a concept which permits both "tatonnement" and "non-tatonnement" properties since it permits any relation between "paper" actions (some of an agent's states may correspond to these) and physical actions.[2]

As far as the critical question of incentives is concerned, one may suppose that the statement of the process already gives effect to them: the two functions which describe an agent's behavior in an economy's process describe the behavior he wishes to engage in. If a "market" process, for example, has better incentive properties than a "command" process that means the actions which agents want to choose, and do choose, in response to signals received may be better actions in the market process, as judged by net performance. But they may also not be better actions: we cannot isolate the effect of incentive differences on the relative performance of the two processes from the effect of other differences between the two processes.

---

[1] The basic elements of this concept of process are taken from the theory of sequential machines [2].

[2] It dispenses, moreover, with the separate specification of "outcome rules" in [3], and "arrangements" in [5].

We can now define a centralized process.[3] To do so consider an additional agent totally outside the economy. Imagine that he observes all the signals transmitted between agents but not the signals from the environment to agents. If with this information alone he can reproduce (on paper) the sequence of actions which the process generates, no matter what the environment sequence, then the process is centralized, or, if one prefers, "centralizable." If one insists on using the term "decentralized" as well it may be defined as equivalent to "not centralized."

This definition has one attractive feature. It suggests rather clearly a class of associated definitions; namely, definitions of the term "less centralized than." Process 1 is less centralized (or "more decentralized") than process 2 if the "difference" between the mythical outside observer's information and the information contained in the agents' states is greater in process 1 than in process 2. The difference can be measured in various ways. One of them is the difference, as between the two processes, in the following quotient: the expected gross performance which the outside observer could achieve (if he were to choose an action sequence for the economy knowing only the signals sent between agents) divided by the performance when the actions are those generated by the process. The more centralized process, in other words, permits the outside observer to choose a better action sequence.

Some previous definitions have failed to meet common discourse in this important respect; while "more centralized" (or "more decentralized") is often used, it has not received careful definition. It is also true, on the other hand, that many persons would think of "less centralized" as meaning, in some sense, "more decisions left to local agents." If more decisions are left to local agents, however, then less information needs to be communicated between agents and the outside observer would know less. But that is precisely what the proposed definition expresses.

Like all formal definitions, this one can serve only as a rough guide to consistent usage when one turns from the precise concept of a process to the complex systems of the real world. Nevertheless, probably most persons would agree that such complex shifts as the Liberman reforms in the Soviet Union, the move toward workers' self-management in Yugoslavia, and the Soviet introduction of the New Economic Policy in the 1920's are all shifts away from centralization in the sense of the definition, and that in the same sense resource allocation in the Soviet Union is more centralized than in the U.S. On the other hand, to decide whether the planned reformed economy of Hungary is more centralized than that of Czechoslovakia is far trickier. Here the definition serves a cautionary purpose.

Empirical Performance Comparisons. Keeping in mind that processes are the objects one ultimately wants to learn about, we turn now to some comparative measurements involving two economies. Two categories of measurement can roughly be distinguished.

In the first category two economies are compared but they are regarded in effect, as two "black boxes"; each box contains a process, information processing facilities to operate the process, and a changing environment to which the process responds. One does not attempt to look inside the box but only to study what emerges; namely, the observed actions of the economies' agents or rather some highly aggregated function of these actions. Foremost in this category are simple growth rate comparisons. One motive for growth rate comparisons is an interest in households' welfare. It is felt, presumably, that if one economy exhibits, say, a rate of growth of per capita output distinctly higher than the other economy, it must then achieve over time a distinctly higher value of some interesting function defined on the future stream of per capita consumptions (including consumption of leisure). Careful studies with this motivation, moreover, would attempt to use aggregate output series for the two economies which give similar weights to the component commodities.

---

[3] The definition follows a suggestion made by A. Camacho.

Since it is so difficult to reduce the many magnitudes entering consumers' welfare to one dimension, another, more impressionistic, kind of comparison has a role to play in assessing the relative performance of the two black boxes; namely, annual disaggregated statements of the bundles purchasable by different classes of income earners in the two economies, together with statements of income distribution. In this spirit is, for example, A. Nove's interesting report of a detailed tour of Moscow shops [8], giving the British and Russian prices of many consumers' goods and taking into account quality where possible.[4]

In a second category of comparative measurement one tries to look inside the black boxes (economies) being compared. One studies economies 1 and 2, which use processes 1 and 2 and seeks evidence that process 1 performs better (according to some common payoff function) in the environment of economy 1 than does process 2 in the environment of economy 2. One might even seek evidence that if process 1 were to be operated in the environment of economy 2, economy 2 would be better off than it is when it uses process 2; such evidence, however weak, is important to reforming economies when they try to imitate existing ones. One might be more ambitious still and seek evidence that superior performance of a process can be attributed to a fundamental property like less centralization.

Perhaps the leading tool currently available for comparative measurements of this kind is the estimation of production functions. Data compel one to study highly aggregative production functions, but imagine for a moment that highly detailed data on the production capabilities of two economies are available. For each economy and each period of time one has a detailed technology matrix, describing the production activities available in each industry during the period, and also a list of capacities for the plant and equipment whose services are inputs entering the activities. Each economy's process in effect computes, after a certain number of steps, the activity levels to be used in the economy during the period (or perhaps only during the part of the period following completion of this computation). The result of operating the activities at these levels will be a new list of capacities available in the next period and an output of goods for consumption in the current period. One process will be a better computer of current activity levels than the other, given the step-completion times made possible by the information-processing facilities placed at its disposal, if it achieves over time a higher value of summed discounted per capita consumption (with the same weights used in computing consumption in both economies).

Suppose, however, that it is known that each process decomposes its computation task into two parts. One group of agents is responsible for computing the current investment decision; i.e., the portion of the capacities available at the start of the period which are to be devoted to capacity expansion. The remaining agents—those responsible for current output—are told this decision. They are then to make good use of the designated portion of current capacities for capacity expansion and of the leftover capacities for production of current consumption goods. We can then judge separately the performance of the "subprocess" operated by the latter group of agents: given a list of capacities available for capacity expansion and a list of capacities available for consumption goods production, one such "current-output" subprocess is better than another if it yields higher current per capita consumption and higher capacity expansion.

---

[4] See also [12].

It is the assumption of a decomposed setup of this sort which permits aggregate production-function comparisons between two economies to serve as clues to the relative performance of their processes. In such comparisons one estimates for each economy the output which accompanies alternative capital-labor combinations and assumes that for a given capital-labor combination differences in output between the two economies are explained by differences in technology and in organization. Moreover, if a multiplicative shift parameter is introduced, the assumption is that shifts in this parameter for an economy reflect "disembodied" changes in technology and organization in the economy. It may or may not be possible crudely to isolate (perhaps by using additional data) the effects of intereconomy technology differences from those of intereconomy organizational differences. Whether it is possible or not, a much stronger interpretation can be placed on differences in estimated production functions if the decomposition just sketched is assumed than if it is not. If the assumption is made, then it is the subprocess determining the use made of given capital-labor combinations that is being judged (perhaps simultaneously with the technology) and not the subprocess determining what particular capital-labor combination is available for use.

A decomposition assumption appears to provide interpretation for several very recent pieces of work. M. Weitzman [13] has estimated a CES production function for the postwar Soviet Union, using the capital stock series furnished by Moorsteen and Powell [7] for 1950–66. He finds that the declining postwar growth rates are very tightly explained by a low elasticity of substitution and a rather modest shift due to technical (and presumably organizational) change. An interpretation: the technology (part of what was called "environment" above) is such that the current-output subprocess is unable to save the economy from the bad effects of a poorly performing investment-choice subprocess; because of low elasticity of substitution, total performance is particularly sensitive to poor choices of total investment (and perhaps investment composition), which result in high capital-labor ratios.

N. Kaplan [4], assuming for the Soviet economy (and for its major sectors taken separately) a Cobb-Douglas production function with constant returns to scale, suggests that the function shifted upward more rapidly in 1951–58 than in 1958–63. His cautiously conjectured interpretation, drastically rephrased in our language, is that the current-output subprocess (whose performance is indicated by the value of the multiplicative shift parameter) permitted good total performance (output growth) in 1951–58 but could not cope with the higher investment rates chosen (by the other subprocess) in 1958–63. Hence one observes low growth in the latter period accompanied by soul-searchings about the current-output process and the beginnings of a reform.

In the case of postwar Yugoslavia one has in approximately 1951–54 a drastic and clearly documented change in the current-output subprocess—a substitution of a less centralized for a more centralized one. The subprocess yielding investment decisions changed, on the other hand, much more gradually. I have reported elsewhere [5] a production-function test of the view that the new current-output process, once established, performed better than the old one. Cobb-Douglas functions were used and the results were inconclusive; CES production functions might be more successful and explorations using them are under way.

What appears as yet unattempted, perhaps because it would be a massive job, is the time series estimation of production functions for both the Soviet and U.S. economies, using comparable series. This would require that for each year in the series the careful effort made by Professor Bergson [1] for one year be repeated. Capital stock and output would be measured, using as weights in both economies the current prices of one of them (and, preferably, repeating the computation with the other economy's current prices as weights). The two labor series would be made comparable by adjusting for differences in the educational composition of the labor force. Using production functions with a multiplicative shift parameter, estimated from such series for each economy, one could ask questions like: given a factor input bundle, which economy produces more output from it, when the shift parameter in each economy's production function takes the value it was estimated to take in some specified year? Is the Soviet economy of 1968 an inferior output producer to the U.S. economy of 19— according to the two estimated production functions (with the shift parameters taking the values of those years)? If 19— is a year in which (on other grounds) the level of U.S. technology is roughly guessed to be that of the Soviet economy in 1968, then performance differences between the 1968 Soviet and 19— U.S. economies might be argued to reflect performance differences in the two economies' current-output subprocesses. One could also ask questions pertaining to the investment subprocess: given the estimated production functions in the two economies, what in each economy is the investment path maximizing the discounted sum of per capita consumptions and how close to such an optimal path is the observed path, actually generated by the economy's investment subprocess?

One major disaggregation in the estimated production functions would—if crude data for it were ever available—allow one to go much further in clarifying the role played by the processes themselves in the observed performance differences between two economies. This disaggregation would be the splitting up of both labor and capital stock into two factors. Labor would be split into the labor used for the operation of the process itself and other labor; capital would be split into the fixed facilities used in the operation of the process and other capital. In both "command" and "market" economies, labor for operating the process would include the central bureaucracy and enterprises' administrative staffs, with the central bureaucracy playing a far larger role in the command economy.[5] Fixed facilities would include data processing installations and communication equipment. With such disaggregated production functions one might hope to have evidence as to whether or not the observed net performance of a command economy, say, could be improved if its process remained the same but more labor and equipment were used to operate it so that each step were completed faster. By allowing, so to speak, each of two processes compared to use the process-operation resources "best" for it (rather than the resources actually used) the investigator would be making a fairer judgment of the two processes.

---

[5] For an experimental effort to measure the amount of information processing that occurs in one industry of a market economy, see [11].

Much would depend, of course, on the number of observations available, and these could perhaps be increased by attempting industry-wide rather than economy-wide studies. Across industries within an economy one might perhaps observe significant variation in the size of process-operation inputs; or possibly a single industry might provide such variation over time.

Short of intereconomy comparisons of such ambitiously disaggregated production functions, one might still obtain useful hints from intereconomy comparisons of crude magnitudes such as "administrative man-hours devoted to computing decisions for the steel industry divided by tons of steel produced." If such magnitudes were much higher in one economy than another, while total capital and nonadministrative labor were roughly the same per ton produced, this would be evidence that one economy's process outperforms the other, at least for some choices of labor and equipment devoted to process operation.

General Conjectures about Decentralization. So far we have chiefly discussed comparative measurements whose ultimate purpose is simply to rank the processes actually used in observed economies. But there are profound general conjectures involving decentralization, and in planning future programs of comparative empirical research one ought to consider what sort of work might contribute to the study of these conjectures.

Perhaps the most pervasive conjectures have to do with "externalities" and decentralization. When the architect of the Czech economic reforms [10] writes that the old (centralized) system was appropriate for "extensive" economic development but cannot cope with the "intensive" development now required, one interpretation is that the old task was simple and the new one complex, involving much more attention to variety and quality in final goods but also many more interactions between the decisions of economic agents—too many interactions to be properly taken account of by a central agency. On the other hand, when people hold that an economy's investment decisions ought not to be left, unguided, to a market (the position taken by French planners, for example), they seem to be appealing to the reverse conjecture. The externalities attached to large investment decisions are in some sense too great to be left to a decentralized market process.

On the theoretical side, certain classic sufficient conditions guarantee that a competitive equilibrium exists and that an optimum is a competitive equilibrium and vice versa; the absence of externalities in production and consumption is one of these conditions. If there are externalities, it seems indeed harder to construct processes which reach Pareto optimality and which are far from total centralization, but it may not be impossible. Some results exist [5] [9] on the way increasing the extent of externalities changes relative performance of some very special centralized and decentralized processes in very simple economies. In these studies performance is measured by the expected summed value, over time, of a specific payoff function and the extent of externalities is measured by the effect of a small change in agent i's action on the marginal contribution to payoff of agent j's action. In some of these results, unfortunately, greater externalities improve the relative performance of decentralization and in others the relative performance of centralization.

A second major conjecture is that the

revolution in information processing technology strengthens the case for centralization in economies. Again the search for simple theoretical results has barely begun. On the empirical side, the disaggregated production functions proposed above would play a role since they might reveal the effect of increasing the process-operation capital on the performance of centralized processes. A complementary line of empirical study would look at what happens to large enterprises (viewed as miniature economies) which have invested heavily in information processing.

A third common conjecture is that in periods of urgency (as reflected, say, in the discount factor that is applied to an economy's stream of payoffs to obtain the performance of the process it uses) the case for centralization is again strengthened. Such periods include wartime mobilization and periods in which a damaged or undeveloped economy is to be rapidly pushed into recuperation or sustained growth.

Conjectures such as these can be abandoned and forgotten, as can the terms "centralized" and "decentralized" themselves. An alternative is to try to clarify and test them and, in fact, to let comparative economics take their resolution as its ultimate goal. But in that case it seems clear that the two groups of investigators briefly dealt with here can hardly hope to progress towards this goal if they are isolated from each other.

REFERENCES

1. A. Bergson, *The Economics of Soviet Planning* (Yale Univ. Press, 1964), Appendix B.
2. J. Hartmanis and R. E. Stearns, *Algebraic Structure Theory of Sequential Machines* (Prentice-Hall, 1966).
3. L. Hurwicz, "Optimality and Informational Efficiency in Resource Allocation Processes," in Arrow, Suppes, and Karlin, eds., *Mathematical Methods in the Social Sciences* (Stanford Univ. Press, 1960).
4. N. Kaplan, "Retardation in Soviet Growth," *Rev. of Econ. and Statis.*, Aug., 1968.
5. T. Marschak, "Computation in Organizations: The Comparison of Price Mechanisms and Other Adjustment Processes," in K. Borch and J. Mossin, eds., *Risk and Uncertainty* (Macmillan Co., 1968).
6. ——, "Centralized versus Decentralized Resource Allocation: The Yugoslav 'Laboratory'," *Q.J.E.*, Nov.. 1968.
7. R. Moorsteen and R. Powell, *The Soviet Capital Stock, 1928–1962* (Richard D. Irwin, Inc., 1966).
8. A. Nove, *Economic Rationality and Soviet Politics, or Was Stalin Really Necessary?* (Praeger, 1964), Chap. 8.
9. R. Radner, "The Evaluation of Information in Organizations," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, 1961).
10. O. Šik, *Economic Planning and Management in Czechoslovakia* (Orbis, Prague, 1965).
11. P. Schmidbauer, "Information and Communications Requirements of the Wheat Market: An Example of a Competitive System," Technical Report No. 21, Center for Research in Management Science (Univ. of California, Berkeley).
12. L. Turgeon, "Levels of Living, Wages, and Prices in the Soviet and United States Economies," in Joint Economic Committee, *Comparisons of the United States and Soviet Economies*, 1959.
13. M. Weitzman, "Soviet Postwar Growth: Some Econometric Aspects," Cowles Foundation Discussion Paper (Oct., 1968).

DISCUSSION

BELA BALASSA: Professor Marschak notes the existence of a gap between two groups of economists interested in the analysis of economic systems: one group attempts to derive conclusions from the study of existing national economies while the other formulates abstract models of economic systems. This cleavage is especially apparent in the centralization-decentralization controversy—the aim of the first group being to evaluate the effectiveness of actual solutions and that of the second to provide definitions of decentralization and to appraise the relative merits of these archetypes.

As an eminent representative of the second school, Professor Hurwicz has put forward several definitions, all of which center on the informational aspects of decentralization. In his earlier writings, he emphasized the characteristics of the transmission of information—including anonymity, self-relevance, and externality—while the definition now proposed is based on the nature of the messages to be transmitted. Under the new definition, the criterion of decentralization is that all messages have the dimensionality of commodity space, as do bids and prices.

I confess a partiality for the definitions used in Hurwicz' earlier work. One reason is that the classification system now proposed depends on what I consider extraneous elements, such as the form of the production function. Thus, an economic process would qualify as decentralized if information on inputs—transmitted by economic agents—permits the central authority to make decisions on outputs and this in turn assumes that the production function is parametric. Accordingly, economic systems possessing otherwise identical characteristics would be considered decentralized or centralized depending on whether the production functions belong to a parametric family.

A further consideration is that the new Hurwicz definition permits directed communication—the transmission of commands on the part of the central authority—which other students of the problem, and Hurwicz himself in his earlier writings, considered as a mark of centralization. At the same time, one can hardly defend this choice on the grounds that "we may be looking for a process that would guarantee optimal resource allocation" since one should properly separate the definitional problem from the welfare implications of the processes it describes. Finally, the definition now proposed takes no account of the amount of information being transmitted. In other words, a system is taken to be decentralized and another centralized even if the former requires more information than the latter, provided that all messages have the dimensionality of commodity space in the first case but not in the second.

If we wish to appraise a definition from the point of view of its applicability to real life situations—and thereby to reduce the gap between the two groups of economists concerned with the analysis of economic systems—there is also the question of whether the definition permits a ranking of alternatives according to the degree of decentralization. The new Hurwicz definition—or for that matter the earlier ones—does not allow for this but distinguishes only two groups of cases. However, in view of the great variety of the solutions we actually observe, our interest is not confined to determining whether an actual case falls into one class or another, but we also ask the question: is A more decentralized than B, or, has A become more decentralized over time?

At any rate, the "more or less" type question is raised in connection with hypothetical cases or ideal (idealized) types of economic systems too—and we have plenty of these from Lange-Lerner to Ward-Domar and from Barone to Kantorovich. With their proliferation, we have a continuum of stages of decentralization and a definition will possess certain advantages if, rather than dividing a continuum into two segments, it enables us to provide a ranking.[1] This is indeed Marschak's declared purpose and he claims to have provided a definition which can be used to indicate degrees of decentralization.

---

[1] This will in general be possible as long as the degree of decentralization is a single-valued function of a given variable. In turn, a ranking of alternatives can permit determining the optimum degree of decentralization.

According to Marschak, a process is centralized if an outside observer can reproduce the sequence of actions the process would generate on the basis of signals transmitted between agents (but not the signals from the environment to agents, such as information on technology). "Process 1 is less centralized (or 'more decentralized') than process 2 if the 'difference' between the mythical outside observer's information . . . contained in the agents' states is greater . . . in process 2." But, as Marschak himself notes in an aside, this definition does not compare degrees of centralization but rather degrees of the centralizability of the process.

Moreover, the welfare implications of alternative solutions again intrude on the scene, inasmuch as Marschak proposes to measure the difference between degrees of centralization as "the difference, as between two processes, in the highest expected gross performance which the outside observer could achieve (if he were to choose an action sequence for the economy knowing only the signals sent between agents) divided by the payoff for when the actions are those generated by the process." At the same time, the definition fails to take account of the possible effects of additional signals. If, for example, gross performance is better in one case than in another but some additional information available to both will change the ranking, should we consider the former to be more centralized than the latter?

A conflict between the Hurwicz and the Marschak definitions should also be noted. We have seen that, in Hurwicz' case, a process is considered decentralized if the central authority can act (presumably optimally) on the basis of messages that have the dimensionality of commodity space. But these are the kind of messages economic agents tend to transmit among themselves, so that such a process would qualify as centralized under Marschak's definition. However we wish to interpret this conflict, the problem remains that both classifications depend on the characteristics of the production function: the existence of parametric production functions appears to be a condition of centralization (or centralizibility) under Marschak's definition just as it is that of decentralization for Hurwicz.

Thus far I have considered the problem of informational efficiency which is at the center of the discussion concerning abstract models. On the other hand, economists dealing with practical situations have stressed the importance of the place of decision making. This difference is more apparent than real as long as we assume the application of definite behavioral rules. Thus, if calculations performed by the central authority faithfully reproduce the market mechanism and the usual assumptions on the absence of increasing returns to scale (including invisibilities) and externalities are fulfilled, gross performance will be identical in the two cases,[2] provided that economic agents carry out the received instructions in the first case and maximize profits in the second. Putting it differently, in models of this sort economic agents are like puppets: they obey external forces and produce the same output with available inputs, irrespective of whether they adjust output to given prices or follow the central authority's decision on the quantities to be produced. However, introducing the problem of incentives, informational decentralization will no longer be identical to the decentralization of decision making and both of these can be of interest.[3]

In examining these two types of decentralization, we can follow Marschak in distinguishing between the allocation of existing resources on the one hand, and the investment program on the other.[4] This scheme of classification is useful in evaluating the actual as well as the optimum degree of informational decentralization and of the decentralization of decision making. Moreover, it permits us to go beyond the oft repeated statement that the market mechanism leads to a Pareto optimum in the absence of increasing returns to scale and externalities. But before going into details, I wish to introduce considerations pertaining to the openness of the economy.

---

[2] This will not, however, be true of net performance, defined as the discounted value of future payoffs, net of information and computational costs.

[3] Parenthetically, Marschak claims that under his definition, the two concepts of decentralization are identical. "If more decisions are left to local agents [he argues] then less information needs to be communicated between agents and the outside observer would know less." This formulation, however, excludes the possibility that information may be transmitted between the agents and the outside observer. Should we take account of such communications, the centralization of decision making may reduce, or even eliminate, the need for the transmission of information among agents.

[4] This is not to say that I would necessarily agree with the empirical solutions proposed by Marschak. Technological change is in part embodied in capital goods and thus is influenced by the choice of investments. This latter choice will also affect the elas-

Since Lange's time, it has become customary to analyze the case of a closed economy and a recent textbook on the organizational alternatives of a socialist economy[5] does not even mention foreign trade. Malinvaud's paper on "Decentralized Procedures for Planning," said to have been "prompted by certain methodological questions arising in connection with French long-term planning,"[6] and practically all contributions to the theory of economic development also assume a closed economy. Yet, as I will try to indicate, our evaluation of alternative systems depends to a large extent on the assumptions made in regard to the openness of the economy.

In considering the case of the closed economy, the theoretical contributions to the centralization-decentralization controversy have taken as their model a hypothetical economy resembling that of the Soviet Union where foreign trade is of marginal importance. However, this model has little applicability to the socialist countries of Eastern Europe, where foreign trade often accounts for 30–40 percent of national income as against 5 percent in the Soviet Union. In fact, rather than being of the closed economy variety, the model applicable to these countries is that of the proverbial small country in international trade theory.

It is hardly surprising, then, that the relationship of decentralization and foreign trade has received much attention in Eastern Europe. The disadvantages of centralized decision making without the use of prices in the allocation process are said to be especially pronounced in the foreign trade sector, in large part because the lack of scarcity prices does not permit a rational choice among exports and imports. It is further maintained that the proposed decentralization of decision making and the application of profit incentives cannot be successful without establishing a price structure linked to world market prices, when foreign trade is necessary both to equate supply and demand at these prices and to avoid the establishment of monopoly positions. This explains that, as Gregory Grossman put it, "to a significant extent *all* of the economic reforms in Eastern Europe—but not in the Soviet Union—have as a main purpose making the given socialist economy more effective as an earner of foreign exchange and as a gainer from the international division of labor."[7] It should be added that the opening of the economy through participation in the Common Market also brought about important changes in the system of planning in France[8] and a consideration of trade opportunities will greatly affect the solutions applied in developing countries, too.

Let me turn now to the centralization versus decentralization issue in the small country case, using the concepts introduced earlier. In regard to the allocation of existing resources, decentralized decision making will now involve information flows emanating from the rest of the world—including bids as well as prices, all of which have the dimensionality of commodity space in Hurwicz' terminology—whereas economic agents in the home country will transmit bids but not prices. In turn, centralization interposes a central authority be-

---

ticity of substitution between labor and capital. At the same time, there is no need to restrict the welfare objective of the allocation of existing resources to the maximization of current consumption as Marschak does; a more appropriate objective is one-period Pareto optimum, including the efficient use of resources in producing investment goods.

[5] Benjamin Ward, *The Socialist Economy*, A Study of Organizational Alternatives (Random House, 1967).

[6] "Activity Analysis in the Theory of Growth and Planning," *Proceedings of Conference held by the International Economic Association*, E. Malinvaud and M. O. L. Bacharach, ed. (St. Martin's Press, 1967), p. 170.

[7] Gregory Grossman, "Foreign Trade of the U.S.S.R.: A Summary Appraisal," in *International Trade and Central Planning*, A. A. Brown and Egon Neuberger, ed. (Univ. of California Press, 1968), p. 341.

[8] Cf. my "Whither French Planning," *Q.J.E.*, Nov., 1965, pp. 537-54

tween foreign and domestic economic units, with its information gathering (from abroad) and data providing (to domestic units) activities in the case of informational centralization, and information gathering, computational, and command transmitting activities in the case of the centralization of decision making.

The choice among these alternatives is by no means certain, since we have to consider informational and computational costs incurred by the central authority and by the economic agents. At the same time, the optimum degree of decentralization will tend to vary from industry to industry and will depend on the size and the structure of individual industries. Thus, the centralization of information may be preferable in the case of atomistic industries and decentralization in others.

The conclusions based on the informational and computational requirements of alternative solutions may be modified if account be taken of incentive effects which are said to favor decentralized as against centralized decision making. At the same time, in the small country model, the objection that profit incentives can lead to results incompatible with a Pareto optimum because of increasing returns to scale, loses its validity as far as traded goods are concerned.[9] I would like to underline this conclusion because I consider the existence of increasing returns to scale as the most important argument against decentralization in a closed economy.[10]

But, how about static external economies and diseconomies? Hurwicz cites the Davis-Whinston proposal, under which a central authority would determine the unit charges to be paid by a firm whose factory emits smoke, to another whose operations suffer from it. He proposes a simplification of this scheme based on joint profit maximization where price setting by a central authority is not necessary. However, aside from the need to compel firms to pay the charges, we face the usual stability problem due to the multiplicity of solutions compatible with joint profit maximization.

At any rate, I doubt the practical importance of static external economies limited to the interaction of two economic units. Even in the smoke or the bee-honey example, whose popularity may indicate the paucity of realistic cases, it can hardly be imagined that only one economic unit would be affected. More realistic cases of external economies and diseconomies are not "directed" but anonymous, in the sense that the person of the recipient is undetermined. The training of workers provides an example of such external economies, and congestion costs of diseconomies. In all instances, a central authority would have to set appropriate payments or charges.

I come now to the question of determining the pattern of new investment. Considerations analogous to those mentioned in connection with the allocation of existing resources hold here, with two additional provisos. First, the information requirement pertains to future rather than to present prices, increasing thereby the cost of information and hence also the economies of information obtainable through informational centralization. Second, we should consider the possible implications for centralization of dynamic external economies, including market-type interactions described by Scitovsky and interactions through income changes discussed by Rosenstein-Rodan and Nurkse. As Hurwicz notes, in a closed economy these could be handled by creating a market for future goods. In an open economy, the answer is different: there is need for information on future prices only, just as in the case when dynamic external economies were absent.

My purpose has been to offer some general remarks on the centralization-decentralization issue in the framework of a small open economy. To simplify the argument I have disregarded considerations such as learning effects, the underestimation (overestimation) of future costs and benefits by economic agents and imperfections in capital markets—all of which may modify the results. The extent of the modification will depend on a variety of factors, chief among them being the level of development, that will have to be taken into account in an investigation of practical cases. Nevertheless, the conclusion remains that the desirable degree of decentralization—whether informational or pertaining to decision making—is greatly influenced by the assumptions made in regard to foreign trade. Students of economic systems are well advised, therefore, to give attention to the implications of the relative openness of the economy for the choice among alternative solutions.

---

[9] This point is well understood in Czechoslovakia and Hungary where the antimonopoly effects of imports in a decentralized socialist economy have been repeatedly noted.

[10] Still, centralization may be necessary in the case of nontraded goods (services) subject to economies of scale, such as electricity, gas, and water.

ABRAM BERGSON: In his thoughtful paper, on which I have been asked to comment, Professor Marschak examines the meaning of centralization and decentralization in respect of economic systems. I for one find the definition of these familiar notions that he elaborates appealing and illuminating. But, as was made clear some time ago I believe by the other principal speaker on this panel, Professor Hurwicz, centralization has two facets. There may be varying degrees of centralization of information. There may also be varying degrees of centralization of authority. Professor Marschak in effect stresses arrangements regarding information, but considers that arrangements regarding authority must be closely related. This must often be so, but not always, and where it is not, I wonder if we would still wish to focus exclusively on information, as Professor Marschak does.

Consider, for example, a capitalist market system where indicative planning has been introduced and where accordingly a vast amount of information beyond what is customary circulates among decision-makers. According to Professor Marschak's terminology, the system in question, I believe, would have to be classified as highly centralized, and in a sense it is. But it seems awkward, nevertheless, to confine oneself to such a usage when from another interesting standpoint, that concerning the locus of decision making, the system is also highly decentralized. It is not difficult to think of other similar cases. What, for example, of a socialist system in which decision making is relatively decentralized, but which is informationally inefficient in the sense that the central planning board collects a great deal of information that it doesn't really need?

In addition to analyzing centralization and decentralization, Professor Marschak explores methods of appraising empirically the comparative performance of alternative systems. I have hardly any fault to find with his brief but penetrating remarks on this matter, but he focuses especially on the use of comparative data on factor productivity. As one who has sought to contribute in this area, I am naturally pleased that Professor Marschak has dwelt on it, but an interesting conceptual limitation of these data to which he refers seems rarely to be fully grasped and may deserve underlining.

I have in mind the fact that such data, as usually compiled, bear on performance only in the use of available supplies of factors and do not really illuminate performance in respect of the creation of such supplies. This is true, not only of data on comparative levels of productivity in different countries at one time, but of data on changes in productivity in a single country over time and for supplies of both different kinds of labor and different kinds of capital goods. Conceivably, for example, an economy might show a high rate of growth of factor productivity and still not be especially effective dynamically because of a failure year after year to produce an appropriate mix of capital goods. Such a failure, of course, might disclose itself in diverse ways, particularly the inordinately limited increase in consumption levels achieved, given the initial and terminal capital stocks, but it could not be detected from the productivity data alone.

Even an avid critic, I think, would find no exception to take to Professor Marschak's concluding conjectures on decentralization. He urges that future progress in the comparative study of economic systems will require the collaboration of two groups of scholars who heretofore have tended to work in isolation: on the one hand, those concerned primarily with empirical inquiry; on the other, those concerned primarily with abstract analysis. His essay serves in effect as evidence of what the latter might contribute to the former. I know I speak for others in either camp, as well as myself, in saying that the evidence is certainly impelling as to the value of such a contribution.

COMPLETED DISSERTATION RESEARCH

MANPOWER SHORTAGES IN LOCAL GOVERNMENT EMPLOYMENT*

By EUGENE J. DEVINE
Simon Fraser University

The phenomenon investigated in this study is the existence of substantial, persistent, and widespread job vacancies or shortages among certain classes of urban government employees. Shortages may be said to exist when an employer seeks to hire additional employees of a given type possessing the same qualifications as those already in his employ and at the same wages, but finds no more available at those wages. Examples of such classes are registered nurses, policemen, teachers, social workers, librarians, and probation officers. Urban governments employ hundreds of different types of employees; yet vacancy problems are confined to relatively few job classes. The Los Angeles city school districts have about 600 job classes, the city of Los Angeles has 900, and the county of Los Angeles has about 1,800. Each of these governments has about 40,000–45,000 employees; yet none reports vacancy or recruitment problems for more than two dozen classes. Why are there shortages for some classes and not for others? And why do governments neither raise wages sufficiently nor extend recruitment activities sufficiently to eliminate the vacancies? Is the observation of vacancies consistent with economic theory?

The nature of the data sought necessitates limiting the scope of the investigation considerably. The paucity of job vacancy data is well known [12], and data on job queues and recruitment costs are even more difficult to obtain. Therefore, I concentrate on three of the shortage classes, namely, nurses, policemen, and teachers, and on the three major local governments in the Los Angeles area for the year 1967. Data for other areas and years are used where available.

Vacancies are related to wages, and methods of wage determination are in turn related to the structure of particular labor markets. An employer confronted with a perfectly elastic supply curve of labor has no problems either of wage determination or of vacancies. He takes the market wage as given and is able to hire as many employees as he wants at that wage. That job vacancies are consistent with equilibrium for a simple (i.e., nondiscriminating) monopsonist was first noted by G. C. Archibald, using the basic theory of wage determination under monopsony developed by Joan Robinson [1] [13]. The equilibrium wage rate and amount of employment for the simple monopsonist are determined by the intersection of the marginal factor cost and value of marginal product curves. The monopsonist would be willing to hire additional workers at the equilibrium wage, but he is unwilling to pay a higher wage because marginal factor cost exceeds the value of the marginal product to the right of the intersection of the curves.

Morton J. Marcus and I have extended this model to incorporate information costs and to show the determination of equilibrium wages, employment, and recruitment expenditures [5]. As an alternative to increasing wages, a simple monopsonist may attempt to shift his labor supply curve by incurring recruitment expenditures, which are analogous to the advertising costs incurred by a seller to shift his demand curve. Recruitment costs refer to costs incurred by an employer to disseminate information about jobs, wages, and working conditions to potential employees. Expenditures on recruitment will enable him to secure a greater number of workers at any given wage rate, and this number will be larger the greater the recruitment expenditures. Average and marginal outlay curves were constructed showing the least-cost combination of wages and recruitment costs to secure any given number of workers. Equilibrium is determined by the intersection of the marginal outlay curve, which is rising, and the value of the marginal product curve. To the right of this intersection, the marginal outlay required to secure additional workers exceeds the value of the marginal product of these workers. The monopsonist would be unwilling to pay higher wages or to incur additional recruitment costs to secure additional workers; yet he would be willing to hire more at the wages currently being paid, and the number of such additional workers he is willing to employ constitute job vacancies.

---

* This paper is based on my doctoral dissertation presented at the Univ. of California, Los Angeles. I gratefully acknowledge the assistance of Werner Z. Hirsch, Armen A. Alchian, and Morton J. Marcus. Financial support was received from the Institute of Government and Public Affairs, U.C.L.A., and the Committee on Urban Economics of Resources for the Future, Inc.

A monopsonist able to engage in perfect discrimination would not experience vacancies in equilibrium. However, even if government employers were able to ascertain the reservation wages of present and potential employees, they are confronted with effective legal constraints against such discrimination.

Following Archibald, I consider monopsony to exist whenever the supply curve of labor to an employing unit rises, whatever the reason. My thesis is that governments are monopsonists in markets for job classes for which they are large employers but not in markets for classes more common to other employers in the community. Monopsony in this sense may be specific to one department within a government or even to one facility within a single department. For example, at the prevailing (common) wage a school district may have an abundant supply of teachers for a school in one neighborhood and vacancies for a school in another. Worker preferences for particular jobs or places of work differ.

The monopsony model implies (and the competitive model does not) that we will observe job vacancies and recruitment costs. But what operational meaning can be attached to these terms? Even a competitive employer will not fill positions instantly, and he must communicate his willingness to hire in some manner. I consider shortages to exist when vacancy rates for a class are persistent and substantial; i.e., when vacancies for a class have been experienced almost continuously over a period of a year or more and when the rates have been double those experienced for the generality of employees. Reliable data on recruitment expenditures by type of employee are nonexistent, but it is possible to observe that the effort and expense undertaken to recruit some types of employees are much greater than for others. Extending the search to distant geographical areas is one indicator of unusual recruitment activity.

Comparable Job Classes

All three major Los Angeles governments are subject to similar legal provisions which require them to pay at least the prevailing wage for jobs comparable to those in the private sector. Section 47 of the County Charter, for example, reads:

> In fixing compensation to be paid to persons under the classified civil service, the Board of Supervisors shall, in each in-

TABLE 1

Job Classes Requiring Continuous Recruitment

City	County	City Schools
Auditor	Accountant	Auditor
Calculating Mach. Opr.	Acct. Auditor	Bookkeeper
Civil Engineering Asst.	Admin. Aid	Clerk
Civil Engr. Draftsman	Admin. Analyst Aid	Custodian
Clerk Stenographer	Admin. Trainee	EDP Prog. Analyst
Clerk Typist	Appraiser I	Gardener
Data Processing Anal.	Asst. Head Nurse	Key Punch Opr.
Duplicating Mach Opr.	Civil Engr. Asst.	Prod. Typist
EDP Programmer	Dep. Probation Off. I	Stenographer
Electrical Engr. Asst.	Deputy Sheriff	Teacher
Gardener Caretaker	EDP Asst. Prog. Anal.	
Jr. Admin. Asst.	EDP Program Analyst	
Key Punch Opr.	EDP Trainee	
Librarian	Grad. Nurse	
Office Trainee	Grp. Supervisor (Probation)	
Planning Asst.	Inter. Stenographer	
Policeman	Inter. Typist Clerk	
Recreation Dir.	Librarian	
Sr. EDP Programmer	Probation Trainee	
Surgical Nurse	Social Worker I	
Telephone Opr.	Stenographer	
Tower Line Mech.	Trans. Typist	
Traffic Engr. Asst.	Typist Clerk	
Utility Accountant		

Sources: Personnel Department, City of Los Angeles; Department of Personnel, County of Los Angeles; Personnel Division, Los Angeles City School Districts.

> stance, provide a salary or wage at least equal to the prevailing salary or wage for the same quality of service rendered to private persons, firms or corporations under similar employment in case such prevailing salary or wage can be ascertained.

Court decisions have established that these provisions are mandatory on the jurisdictions, although a determination of the prevailing wage in good faith is not subject to judicial review [14, pp. 22–23, 95]. Prevailing wages are ascertained with the aid of a joint survey of wage rates in the area which is conducted annually. Each jurisdiction makes general wage adjustments annually, if warranted, and specific adjustments may be made during the year.

The recruiting experiences of the three governments have been broadly similar. Table 1 shows jobs for which the governments find it necessary to conduct continuous recruiting and for which vacancies usually exist. Vacancy rates are not necessarily high for all these classes; continuous recruitment is necessary for some because of high turnover. All governments reported continuous recruiting for typists, stenographers, auditors, and data processing analysts and programmers. Several positions are listed which the governments do not have in common. The city and county both list civil engineering assistants and policemen (deputy sheriffs), which have no counterparts in the city schools. The city and county both list nurses and librarians. The city schools also employ these classes, but experience no vacancy problems; both are paid on the same schedule as certificated personnel (teachers), and these salaries are substan-

TABLE 2

Job Classes With Queues

City	County*	City Schools
Carpenter	Clerk (Hvy. Lifting)	Cafeteria Wkr.†
Clerk	Clerk (Permanent)	Carpenter
Comm. Relief Clerk	Tour Guide	Clerk ‡
EDP Trainee I	Med. Case Worker	Painter
EDP Trainee II	Appraiser Trainee	Stenographer‡
Fireman	Custodian	
Maint. Const. Helper	Fire Contrl. Laborer	
Meter Reader	Guard I	
Painter	Laboratory Asst.	
Park Service Attnd.	Laboratory Attn.	
Plumber	Telephone Opr.	
Policewoman	Fireman	
Real Est. Appraiser		
Refuse Collect. Fmn.		
Student Worker (Summer)		
Student Prof. Wkr. (Summer)		
Truck Operator		
Welder		

* Sample classes only.

† Downtown and certain outlying areas only.

‡ San Fernando Valley only.

Sources: Personnel Department, City of Los Angeles; Department of Personnel, County of Los Angeles; Personnel Division, Los Angeles City School Districts.

tially higher than those paid by the city and county.

Table 2 shows job classes for which queuing exists in the sense of long waiting lists or large numbers of applicants per job opening. It is more difficult to make comparisons here because the county's list is not complete. Firemen are listed by both the city and county. Policewomen appear on the city's list, and the county is known to experience no difficulty in recruiting female deputy sheriffs. The city reports queuing for seven craft job classes and the city schools for two. The county also experiences queuing for these classes, although none are included on the sample list. All jurisdictions follow the practice of paying craft positions at rates 11 percent below those specified in the agreement between the local contractors' association and the building trades unions. To the extent that building trades unions have been successful in restricting entry, one would expect these rates to be above market-clearing rates. Moreover, the rates specified in the agreement are for construction work, whereas government craftsmen are employed in maintenance work. Construction rates in the private sector are much higher than maintenance rates, as indicated in the wage surveys.

Registered Nurses

Shortages of registered nurses are characteristic of most labor markets in the United States, but they are largely confined to the hospital nurse segment of the market. Hospitals and other institutions employ about 63 percent of all employed registered nurses, and vacancy rates in large, nonfederal general hospitals averaged about 20 percent of budgeted positions in the early 1960's. Rates were high for both private and public hospitals, though somewhat higher for public hospitals. These vacancies exist even though an estimated 33 percent of all registered nurses are not employed in nursing. Vacancy rates are much lower and salaries are higher among such other types of nurse employers as school boards, public health agencies, and industry. Hospital labor markets in most large areas are characterized by monopsonistic agreements among members of local hospital councils, and there is evidence that such collusion has been successful in depressing wage rates.[1]

Registered nurses are the only one of the three shortage occupations discussed for which there is an appreciable number of comparable jobs in the private sector and which is subject to prevailing wage provisions. In fact, the county pays its hospital nurses about 10 percent more than private hospitals; yet its highest vacancy rates are for this class. Nurses' wage rates are set by collusive agreement among private hospitals in Los Angeles. Most private hospitals belong to the Hospital Council of Southern California, which establishes wage rates based on the recommendations of the personnel consulting firm of Griffenhagen-Kroeger, Inc. [10]. In a 1962 report to the council, this firm urged members to subscribe formally to an apparently effective unwritten agreement to avoid attempting to hire employees of other members, and to refrain from offering better salaries, benefits, or more rapid advancement to employees seeking to change jobs [6, p. 7].

The County Department of Hospitals had about 2,000 authorized positions for registered nurses at its nine facilities in 1967, more than half of which were at County General Hospital. Approximately 1,500 of these positions were nonsupervisory. Of 572 positions for graduate nurses (beginning registered nurses with no experience), 172 were normally vacant, and 234 of 944 positions for assistant head nurses were usually vacant. Thus, the vacancy rate was about 27 percent for these positions and about 20 percent when related to all nurse positions in the department. Private hospitals in Los Angeles were known to have vacancy problems, although the rates were lower than for the county. Positions at County General Hospital, in particular, were rendered less attractive because of high vacancies and heavy workloads. The county engages in nationwide recruiting for hospital nurses, mainly through advertising in out-of-state newspapers and in professional journals.

The County Health Department had about 650 authorized nurse positions in 1967, 200 of which were for public health clinic nurses and 360 for staff public health nurses. The first class consists of graduate nurses with no special training, and the salary range is the same as in the county hospitals. Public health nurses have specialized training, and their salaries are about 5–6 percent higher than those of assistant head nurses in the hospitals. These nurses work in about fifteen health districts throughout the county. There are usually no vacancies in any of these positions, and turnover is low.

The city operates a Receiving Hospital at which it employs 74 surgical nurses. Continuous recruitment is undertaken for this class, but the vacancy rate is well below that for the County Department of Hospitals. The city schools employed 320 registered nurses in the fall of 1967. They draw on a smaller pool than the hospitals because these nurses are required to have bachelor's degrees. Nurses are paid on the same salary schedule as teachers, and these salaries are much higher than for hospital nursing. There were no vacancies in late 1967, and the city schools report low turnover and no difficulty recruiting these nurses.

---

[1] This paragraph is based on an intensive study of nurses' labor markets in the United States recently completed by Donald E. Yett [15]. My findings for the Los Angeles market are in close agreement with his. Yett also uses a monopsony model.

Policemen

Police departments in 90 percent of cities with populations over 100,000 were below authorized strength in 1966–67, and the weighted average vacancy rate was about 5 percent [3] [7]. Almost all vacancies are in the patrolmen ranks, and most reported rates would be about 25–50 percent higher if related to these ranks alone. About 70 percent of all cities pay policemen and firemen the same salaries, although shortages of firemen are unknown. There has been some movement away from such "parity" in response to police recruiting difficulties, but this is vigorously and often successfully opposed by firemen's organizations, while police groups support it.[2]

Police applicants commonly need no experience and only high school graduation. However, various physical, mental, and character qualifications reduce the potential number of policemen greatly. According to one estimate, only about 14 percent of California males in the relevant age group would be able to meet them [4]. It is not clear when the police vacancy problem originated, but there is some evidence that it began in the early 1960's, coinciding with declining unemployment rates and increased military mobilization.

The city of Los Angeles has a separate salary ordinance for policemen and firemen which embodies the parity principle.[3] The county follows the city and adopts identical rates for its police classes. In 1967 the county normally had about 150 vacancies for 2,650 deputy sheriff positions, or about 5.5 percent. In the spring of 1967 the city had 5,400 authorized police positions, 3,800 of which were in the patrolmen ranks, and about 200 vacancies, or 5 percent of patrolmen ranks. Citing the police recruitment problem, the mayor proposed a 5.5 percent raise for policemen and 2.75 percent for firemen. He stated that it would cost an additional $1,000,000 if firemen also received 5.5 percent. The firemen opposed this, and the city council voted a 5.5 percent raise for firemen. The mayor vetoed this, but the council overrode the veto. Both the city and the county reported queuing for firemen's jobs. Exams are given infrequently, sizable eligibility lists are built up when they are given, and there are sizable lists of those waiting to be notified of future exams. At the prevailing salaries, the city was able to fill about 95 percent of its patrolmen positions and had queues for firemen. Parity compounds monopsony, for in attempting to attract 200 additional policemen by raising wages, the city would have to increase the salaries of not only the 5,200 policemen of all ranks already employed but also of 3,400 firemen.

Police recruitment costs are much greater than for most other classes for both the city and the county, and both recruit continuously. Both jurisdictions recruit out of the area, and the city from time to time recruits in distant parts of the country.

Police vacancy rates are generally higher than overall vacancy rates, which are normally about 2–3 percent for the city and county. Increased use of civilians in police departments has been one response to the situation, and these vacancy rates also are lower than for sworn personnel in most cases. The city had about

---

[2] The best study of police compensation and the parity issue I have found is one by William F. Danielson, Director of Personnel for the city of Berkeley [4].

[3] Ordinance No. 89,935, Salary Standardization Ordinance for Firemen and Policemen of the City of Los Angeles.

1,600 such positions in September, 1967, and a vacancy rate of about 3 percent. There were no police vacancies in this month, but this was regarded as temporary. The county had a vacancy rate of about 4 percent for 1,800 civilian positions in the Sheriff's Department. Some additional data are available from a survey of 36 cities with populations of 300,000 to 1,000,000 conducted in 1967. The median vacancy rates for policemen were 4 percent for all ranks and 5 percent for patrolmen ranks, and the median for civilian employees in these departments was 2½ percent. Vacancy rates were lower for civilians than for all police ranks in 23 cities, higher in 9, and equal in 4 [8].

The city had 94 policewomen positions and the county had 200 in 1967. Policewomen receive the same salaries as men, and there was queuing for these positions in both jurisdictions. The city had no difficulty in filling 357 police motorcycle assignments, in spite of the fact that the rate for injuries on duty in this assignment is more than double the departmental rate. Motorcycle officers receive a $77 per month bonus, which it is not necessary to pay to all other officers.

Teachers

The central features of teacher shortages in Los Angeles and most other cities are: (1) they are concentrated in certain subjects and neighborhoods; (2) salary distinctions are not made on the basis of subject or location, but on the basis of years of schooling and service; (3) shortages do not generally take the form of unfilled positions, but rather of inadequately qualified teachers in these positions [9].

The Los Angeles city schools have difficulty in recruiting qualified high school teachers of mathematics, science, industrial arts, and women's and men's physical education. This is consistent with the experience in many other areas, except for the last field, in which there is usually an abundance of qualified applicants. Difficulties in recruiting men's physical education teachers are recent and are attributed to a state credential reform which makes physical education majors ineligible for promotion to administrative positions. Shortages in the sense of inadequately qualified teachers are heavily concentrated in the ghetto areas of south and central Los Angeles. There is a surplus of qualified applicants for schools in West Los Angeles and the San Fernando Valley.

In a 1965 study, Dr. Kenneth A. Martyn compared the percentages of substitute teachers and those with provisional credentials, which require only three years of college, in schools in four disadvantaged areas with those in three privileged areas. In the elementary schools these percentages were not much higher in the disadvantaged areas except for Watts, where about 6.6 percent of the teachers were in these categories compared with 3.3 percent in the privileged areas. In the secondary schools these teachers constituted a much higher proportion of the faculty in the disadvantaged areas; in Watts the percentage was about 17 percent, compared with less than 3 percent in the privileged areas [11]. Prior to the start of the 1966–67 school year, the State Board of Education authorized 239 school districts to hire a total of 988 teachers with new provisional credentials. The city schools, which employ about 15 percent of the teachers in the state, were authorized to hire 331, or one-third, of these teachers [2, p. 82].

The city schools send recruitment teams all over the country to recruit teachers. The only other class for which nationwide recruiting is undertaken is top-level administrators. The marginal costs of recruitment rise sharply as the number of out-of-state teachers the city schools seek to hire increases because recruiters must visit increasingly smaller cities and institutions and incur increasingly greater time and travel costs for each person hired.

Conclusions

The Los Angeles governments appear to be competitive buyers in markets for jobs which are common in the community, or at least to be constrained to pay competitive wages by legal provisions. These jobs are a majority of the job classes in government. There are situations in which the prevailing wage policy results in wage rates which are higher or lower than necessary to attract the number of personnel effectively demanded. The first situation occurs when the labor market is characterized by entry restrictions, as in the building trades, and the second when it is characterized by monopsony among private employers, as in hospital nursing.

Vacancy problems are common for occupations for which government agencies are large employers. Substantial and persistent vacancies and heavy recruitment costs are consistent with the monopsony model, but not with the competitive model. They may, of course, be consistent with a number of other hypotheses. Monopsony, in the sense of a positively sloped labor supply curve, may be specific to one department within a government or even to one facility within a department. It is compounded by the common practice of paying the same wage rates for jobs for which employee preferences differ.

REFERENCES

1. G. C. ARCHIBALD, "The Factor Gap and the Level of Wages," *Econ. Record,* Nov., 1954.
2. ARTHUR D. LITTLE, INC., *Teacher Supply and Demand in California, 1965–1975,* California State Department of Education, Sacramento, 1967.
3. RAYMOND L. BANCROFT, "Municipal Law Enforcement, 1966," *Nation's Cities,* Vol. 4, Feb., 1966.
4. WILLIAM F. DANIELSON, *Police Compensation,* The President's Commission on Law Enforcement and Administration of Justice (Washington, 1967).
5. E. J. DEVINE AND MORTON J. MARCUS, "Monopsony, Recruitment Costs and Job Vacancies," *Western Econ. J.,* Sept., 1967.
6. GRIFFENHAGEN-KROEGER, INC., *A Report Presenting Recommendations for Hospital Salary Administration* (Los Angeles, July, 1962).
7. INTERNATIONAL CITY MANAGERS' ASSO., *The Municipal Year Book* (Chicago, 1967).
8. KANSAS CITY MISSOURI POLICE DEPARTMENT, "1967 Survey of Municipal Police Departments" (Kansas City, Mo., Oct., 1967).
9. JOSEPH A. KERSHAW AND ROLAND N. MCKEAN, *Teacher Shortages and Salary Schedules* (McGraw-Hill, 1962).
10. *Los Angeles Times,* July 27, 1966.
11. KENNETH A. MARTYN, *Report on Education to the Governor's Commission on the Los Angeles Riots* (Los Angeles, Nov., 1965).
12. N.B.E.R., *The Measurement and Interpretation of Job Vacancies* (New York, 1966).
13. JOAN ROBINSON, *The Economics of Imperfect Competition* (Macmillan, London, 1933).
14. TOWN HALL, *Pay Policies for Public Personnel* (Los Angeles, May, 1961).
15. DONALD E. YETT, "An Economic Analysis of the Hospital Nursing Shortage" (unpublished Ph.D. dissertation, Univ. of California, Berkeley, 1968).

RESOURCE ALLOCATION WITH PROBABILISTIC INDIVIDUAL PREFERENCES*

By RICHARD ZECKHAUSER
Harvard University

An individual will rarely be able to determine exactly his future level of demand for a good. This demand will depend upon the evolution of his preferences, something that he may not be able to predict accurately. This need not reflect an inability to know his own future mind. Rather, events over which the individual has little control may significantly affect his future preferences. An individual would have to know whether his house would catch fire to predict his demand for fire-fighting services. He would have to know whether he would be involved in an accident before he could estimate his need for a hospital accident emergency room.

In normal circumstances, the inability of individuals to predict their future demands for goods is not of great consequence. A good is supplied to the market and is consumed according to the individuals' preferences at that time. Only if action must be taken at present to ensure an efficient level of provision for the good is it important to make accurate predictions. Thus, for example, if there is a significant lead time for production, it will be most useful to be able to predict aggregate demand with some accuracy.

This essay considers situations in which individuals who are unable to predict their individual portions of future total demand may find it useful to make some decisions on the provision of a good before preferences become known. For the most part it is concerned with goods which by their very nature make accurate prediction of future preferences impossible. At best, for such goods, it will be possible to attach some probability assessment to possible sets of preferences for the individual. The acronym PIP, derived from probabilistic individual preferences, is used to refer to the goods of this nature. Most frequently, the probabilities attached to these sets of preferences will be determined subjectively rather than objectively. For the purposes of this essay this distinction is not of importance.

The Efficient Provision of PIP Goods

Option Values. Consider a community that must decide whether to supply itself with a PIP good, perhaps a supply of plasma at a blood bank, or a fire engine to protect its homes. Following conventional efficiency dictates, the community should estimate future levels of demand for the good and decide whether perfect price discrimination against future consumers would yield sufficient returns to enable discounted revenues to cover discounted expected costs. If the decision is negative, we are told that the community should not provide itself with the good. This stipulation holds whether or not consumption by one individual reduces the amount available to be consumed by another, whether or not the good has any aspects of a collective-consumption good.

Many PIP goods serve a preventive purpose; for example, programs to deter crime or to immunize against disease. If such goods are provided, it may never be known who are the major beneficiaries, who would have been the victims of the crime or the sufferers from the disease. For such goods, discriminatory pricing questions could only be asked on a hypothetical basis. If you have been identified as a future target of a mugging or a future victim of polio, how much would you pay to prevent the unfortunate happening from taking place?

---

* This paper was condensed from Essay I of my dissertation, "Studies in Interdependence." Robert Edelstein, Emmett Keeler, and Professors Howard Raiffa and Thomas Schelling made helpful comments.

In an article dealing with commodities "the purchase of which is infrequent and uncertain," commodities which are in effect PIP goods, Burton A. Weisbrod argued that even if there will be perfect price discrimination, the option values of potential consumers would influence the decision whether such commodities should be provided. He stated, "If these consumers behave as 'economic men,' they will be willing to pay something for the option to consume the commodity in the future."[1]

The key question for Weisbrod, then, is what will be the magnitude of these option values. The amount that an "economic man" will pay for the right to consume a commodity in the future will depend upon the price that he will be charged for the commodity. If he will be faced with perfect price discrimination, should he be interested in consuming the good, he will be indifferent whether or not he purchases it. He will not be able to reap any consumer's surplus at the time of purchase; the availability of the good will in no way be able to increase his welfare. His option value to keep the purchase available will be zero. I disagree with Weisbrod and assert that in this regard the conventional efficiency dictate is not called into question.

Risk Spreading. The option value, discriminatory price approach to the allocation decision problem is not sufficiently rich. It overlooks some important considerations, one of which is the risk-spreading property of PIP goods. Every potential consumer of a PIP good is in an uncertain situation. He does not know whether he will wish to consume the good in the future. The value in use of many of these goods can be very high.[2] If the actual consumer is to be faced with perfect price discrimination, then each potential consumer is in the position of having some probability of incurring a substantial charge.

However, it may not be necessary to charge the actual consumers even a small fraction of their discriminable price. The community of potential consumers as a whole can agree to share the costs of making the good available, with little or no additional cost to be borne by the actual consumers. In this way, each individual will bear a small certain cost rather than run the risk of a much larger but considerably less probable cost. If the individuals in the community are risk-averters, this indirect form of risk-sharing may be of value to them.

The possibility of risk-sharing introduces a new consideration to the question as to whether the good should be provided. Let me illustrate with the fire engine example. Consider a community of one hundred identical individuals, with identical utility functions for wealth, log (w), each with a house they value at \$10,000, and each with non-house wealth of \$10,000. Every year one house, no more, no less, burns down. The community must decide whether to get a super fire engine that stops all fires instantaneously. The rental on the fire engine is \$12,000 per year in comparison to the \$10,000 that the individual whose house starts burning would pay for the services of the engine. (If the income effect for housing were positive, this would somewhat reduce the \$10,000 figure, and conversely.)[8] Following conventional efficiency dictates, the community would compare the amount that could be secured for the good through perfect price discrimination with the cost of provision. It would decide not to rent the engine.

---

[1] "Collective-Consumption Services of Individual Consumption Goods," *Q.J.E.*, Aug., 1964, p. 472.

[2] This is particularly true of goods in this class which are primarily used to avoid or ameliorate unfortunate occurrences. Many goods with such a purpose fall into the class of PIP goods. The expected time until a highly unlikely loss is incurred is great. All losses or expenses in this essay are assumed to be discounted. With uncertain occurrence time, the discounting process is complicated.

However, the individuals within the community are risk-averters. They would be willing to pay a premium to reduce the variance in their payoffs. If they share the costs of providing the engine, they can accomplish this. Let x represent the certainty equivalent dollar value for the lottery that gives the individual's payoff in the absence of the fire engine. We have, $\log(x) = .99 \cdot \log(20{,}000) + .01 \cdot \log(10{,}000)$. This gives, $x = 19{,}861.85$, and $20{,}000 - x = 138.15$. To avoid one chance in one hundred of a \$10,000 loss, each individual in the community would be willing to pay \$138.15. This amount can be looked at as the individual's option value for the right to consume at zero cost the fire engine's services at any time during a one-year period. The sum of the individual's option values would be \$13,815, an amount well in excess of the rental price.

What, then, of the decision not to rent the engine? Each individual would rather pay his share of the rental rather than be faced by the lottery that confronts him if the fire engine is not rented. If the only alternative to renting the engine is to let the houses burn and have individuals suffer the losses, then the engine should be rented and the rental cost shared equally. This will involve a loss in efficiency with regard to resource use, but there will be a more than compensating gain, because, unlike the costs of a fire, the costs of fire control can be spread among all the members of the community.

But another, more attractive alternative may be available. It may be possible to spread the cost of the fires through an insurance pool in the community. If each individual agreed to pay one-hundredth of all fire losses in the year, his annual cost would be \$100. This would give him a considerable saving over paying his share of the engine rental cost. If there are no institutional constraints to prevent the operation of an insurance plan, then so long as administrative costs are not prohibitive (below \$2,000 in this example), this alternative will be the most attractive.

The Community Insurance Decision When the Size of Loss Is Variable. The assumption that a single house will burn each year is certainly unrealistic. There may be major conflagrations, and even if house burnings are independent events, there will be some variance in the number of houses burned each year. The fire engine has the advantage that it eliminates all such variance in total cost to the community. This improves its attractiveness relative to an insurance scheme. If there is significant variance in the number of houses that burn each year, the community might choose the fire engine over the community-wide insurance plan even though the latter would have a lower expected cost per individual.

The limiting case of complete contagion would have the only possible fires in the community destroy all houses, the annual probability of such a catastrophe being one in one hundred. In this case the insurance plan would serve no risk-spreading purpose. The lottery for each individual's payoff would be the same as the one in the absence of insurance. Following the analysis above, the fire engine should be rented.

---

[8] Income-type effects can be important if PIP goods are of exceptional value in use. For example, Condition X has a .8 chance of being fatal. The discriminable amount an X-sufferer would pay for Treatment B which would reduce this probability to 0 might not be much greater than that for Treatment A (assuming B unavailable) which would merely reduce it to .4. If there were one unknown sufferer in a large community, the result would be much different. The total discriminable payments from all members to make Treatment A available to the unfortunate member would be approximately twice that for Treatment B. Where specific goods, such as houses, make up the potential losses to be prevented or insured, the amount that would be paid at the higher, before-loss income is the relevant amount for efficiency calculations.

Assume (Virginia) that there were no fire engines. We observed that a risk-spreading plan in a community will be the more effective the less positive correlation there is between significant losses for its individual members. This observation may be of import if each individual will be charged a proportional premium above his share of community losses if an insurance plan is instituted, if, for example, there are administrative costs that are some fixed percentage of total reimbursed losses. The individuals might find it desirable to establish an insurance plan that would reimburse losses if one or two houses were lost in a year but not if a great many had burned.

By way of illustration, assume that there are but four possibilities. The number of houses that burn in the community with the probabilities of occurrence are: 0, .75; 1, .15; 10, .07; 50, .003. The expected number of houses burned is one. Our community would find it desirable to reimburse losses from single house fires if the premium over fair actuarial value were not more than 38.1 percent. It would be worthwhile to insure losses in ten house fires as well if the premium were less than 33.9 percent. Only if the premium fell below 17.2 percent would the community wish to include losses from fifty house fires in its insurance plan.[4]

Utilities of Anticipation Versus Utilities of Use

There are some special characteristics of goods the future consumers of which are unknown that render the actuarial model of decision inadequate. The lottery model is essentially static. The decision-maker is assumed to act as if the probabilistic outcome is to be determined immediately when his decision is made. Pleasure or pain derived from anticipation or anxiety over the outcome of the lottery does not enter into the model. But these feelings are not imaginary. We cannot afford to neglect these real and potentially significant utilities of sensation. They must be considered in addition to the familiar (what might be called) utilities of use.

Many individuals would gladly incur significant costs so that they could live in a community with a museum or symphony even though they are unlikely to make use of these cultural advantages. A shopper might pay to have a wide selection of goods available even though she knows she will probably choose her purchases from among a very few. We have in effect what might be called a "utility of variety." Consumers like and get positive utility from the fact that goods are kept available. From what we can observe, the amount that these individuals would pay to keep these facilities available is well in excess of the expected value to them, the probability they use them times the difference between the price they will be charged and the discriminatory price that they could be charged. For most individuals the discriminatory price would be low enough so that risk considerations do not play a major role. It would seem that another factor is present. It is what we might think of as the utility of knowing you can use something even though it is unlikely that you will use it. It is a utility that is defined independently of the use itself.[5]

PIP Goods, Insurance, and Markets for Contingent Claims

Prospective and Retrospective Benefits. In our discussion thus far we have found a number of relationships and similarities between insurance mechanisms and PIP goods. An insurance policy gives one claim to future resources should some prespecified event occur. One takes out the policy because it provides an expected positive benefit. It cannot be known in advance whether one will actually benefit from having the insurance policy. Generally, in retrospect, one will not prove to have benefited. The events that we insure against usually have a low probability of occurrence. When one shares in the provision of a PIP good so as to be included as a potential consumer, one is in a similar position. There may only be a small chance of benefiting from the consumption of the good, but the benefit may be sufficient to make the premium (sharing in its provision) worthwhile.

---

[4] Omitted section: "The Individual Insurance Decision when the Size of Loss is Variable."

[5] Omitted section: "Misestimation of Probabilities."

Quite frequently, PIP goods are provided by a political unit on a nonvoluntary contribution basis. In insurance, the equivalent arrangement is some form of mandatory policy; for example, social security. This provision, whether for PIP goods or insurance, can be justified on efficiency grounds only if the gains of some individuals, as measured by the excess of expected consumer's surplus over share of provision cost, more than outweighs the losses of those who would not join voluntarily.

An insurance policy may provide a nonactuarial benefit that in some sense is parallel to the utility of variety that is associated with PIP goods. It is a utility that is not directly connected with the use of the goods or the money that is received should the insured against event occur. It is the pleasurable sensation that is derived from knowing that one has protected one's self (or one's family) in case an unfortunate contingency should arise.

Thomas Schelling kindly suggested the following demonstration of the value of this "peace of mind," what we might call a utility of security. A man with a medical ailment is fully aware of the seriousness of his condition but does not know whether it will exclude him from life insurance. He asks his wife to find out and to pay the premium if his life can be insured. We would likely agree that she may be doing something useful if she can deceive her husband should he be uninsurable and convince him that his life is insured.

Insurance and Markets for Contingent Claims. When one takes out an insurance policy, one sacrifices current assets in exchange for a contingent claim to be paid, if at all, in the future. An insurance policy represents a trade in a contingent claims market that is restricted in two ways: (1) In an open contingent claims market, future as well as present payments can be used to purchase contingent claims. This restriction will not be of consequence if there are perfect capital markets and if borrowing and lending rates are identical. (2) In an unrestricted contingent claims market, one can exchange a contingent as well as a certain claim for a contingent claim.

If one wished to take out medical disability insurance rather than pay some fixed amount, it might be far preferable to agree to pay an appropriate fixed percentage of future income that would result in the same expected payment. Carrying this reasoning one step further, it would likely be even better if one could work out an arrangement under which the percentage was higher the higher was future income.

There may be situations in which contingent markets or insurance are not available but where good substitutes can be found. A farmer who cannot find a futures market in the particular commodity which he produces can in effect insure his price if he can sell short a related or derivative commodity which will have parallel price movements. In financial markets it is often possible to secure a form of insurance through arbitrage-type operations; for example, one can sell short in a stock for which one holds convertible bonds. Some situations call for more inventive action. An individual who cannot buy storm insurance on his home in a hurricane belt might cover himself by investing in a roof repair company. In all these situations, the object is to diversify. The extensive literature on the subject of diversification is of consequence because perfect insurance and contingent claims markets do not exist. To the extent that balanced portfolios can be secured, a suitable substitute for these markets can be found.[6]

The Collective Provision of PIP Goods

There are at least three factors that might make it best to provide PIP goods on a collective basis: (1) If the probability is small that any one individual will consume a good in a given short period of time and if it must be kept continually available, there will be considerable waste if each individual provides for himself. We find it efficient to store plasma in a common pool in blood banks rather than have each individual keep a personal supply. (2) PIP goods possess aspects of collective-consumption goods and for the conventional reasons are best provided on a collective basis. For example, the consumption of the services of a fire engine by one individual in no way reduces the amount that is available to be consumed by another.[7] An essential collective-consumption element of some PIP goods is that it is desirable to keep an inventory available. Thus, in our plasma example, the good itself is of the private consumption variety, but the existence of a stockpile upon which any member can draw has collective-consumption aspects, consumption of the "good" by one individual does not reduce the amount available to be consumed by another. (3) If PIP goods are provided on a collective basis, the costs of provision can be shared, and the as yet unknown consumers need not be charged any great amounts. In this way some useful risk-spreading can be accomplished.

On an equity basis, we might object to the collective provision of collective-consumption goods that are not enjoyed equally by all contributors (assuming equal incomes and contributions to provision). It might seem that such an objection would apply with a vengeance to PIP goods, but such is not the case. Collective provision of PIP goods can only be called inequitable if on a prospective basis it is clear that some community members will have a higher expected benefit. Even then, there might be disagreement.

Final judgment may depend whether higher probabilities of use of higher values in use are the primary explanation of higher expected benefits for some individuals. If the former, it might be relevant to know why the probabilities are higher, whether they depended upon conscious choices. On an equity basis, we would be more likely to object to a subsidized public facility to treat those injured skiing than to a subsidized public facility to treat those who suffer from a particular inherited illness, even though we may be better able to predict who will be the beneficiaries of the latter.

The lessons of equity considerations in situations of uncertainty have not been thoroughly enumerated. These lessons may be applicable to the most basic of philosophical considerations. If we state that society should establish a mechanism

---

[6] Omitted major section: "Inefficiency and Restrictions on Contingent Claims Markets."

[7] In the unlikely event that the demand periods for the services overlap, there will be the congestion problem that is frequently associated with "nearly collective goods."

EFFECTS OF TAXATION ON RISK-TAKING*

By AIKO N. SHIBATA
University of York, England

Introduction

This paper examines the effects of the general sales tax on the distribution of the economy's total capital between risky and sure industries.

The effect of taxation on investment has been subject to a good deal of discussion in the literature of public finance [1] [4] [5] [6] [9] but most of the analyses have been confined to studying the effects of income taxation on the composition of an individual investor's portfolio.

Few studies exist which analyze the relationships between the changes in the investors' choice and the changes in allocation of real capital among different risk ventures.

This paper represents a modest attempt to construct a new model capable of analyzing the allocation processes of real capital among industrial sectors with different degrees of risk, assuming that the investors evaluate the investment returns in terms of the mean and standard deviations. Having constructed such a model, the paper analyzes the effects of the general sales tax on the distribution of real capital among different industries.

The study concludes that the general sales tax causes a shift of real capital out of sure industries into risky industries. The conclusion is, of course, based on rather stringent assumptions made and has to be interpreted with care, but may be a useful starting point of further discussions of the problem in question.

The Model

Let us consider an economy consisting of production and consumption of three goods, X, Y and Z. The consumption of Z remains unchanged from period to period, while consumption of X and Y fluctuates. However, the gross income of the economy is stable so that total expenditure on the three products combined is unchanged over the periods concerned. Let us denote this total expenditure C, and assume that $(1-a)$ fraction of C is spent on goods Z and the rest (i.e., Ca) is distributed randomly between spendings on X and Y. By using a stochastic variable, b, we say that b fraction of expenditure on non-Z goods is expenditure on X and $(1-b)$ fraction is expenditure on Y.

On the production side, we assume that there exists a time lag between investment and production; namely, to meet the demand of the current period, investment must be carried out in the preceding period. This is a typical short-run case. Once the investment is completed, producers cannot change their capital facilities to meet the current demand even though it differs from that anticipated, and the actual rate of return of capital invested in one industry turns out to be different from that of the other. But the investors need not make an advance commitment about the amount of labor they hire at the current period: labor is assumed to be instantaneously adjustable to the current demand conditions. In a competitive labor market, assumed for our economy, the actual wages, P_L, are the same in three industries producing X, Y, and Z, and production in each industry will be carried to the point where the value of the marginal product of labor is just equal to wage, P_L, paid by the firm for the service of the labor. Let the production functions in three industries be of the Cobb-Douglas type, namely:

---

* This paper is based in part on my doctoral dissertation submitted to Columbia University in 1968. The author wishes to acknowledge her indebtedness to Professors C. S. Shoup and K. J. Lancaster for their research supervision, and H. Shibata, Y. Fukazawa, N. Hashimoto, and all members of the Staff Seminar at the Univ. of York, England, for their comments.

$$X = L_x^{(1-\alpha)} K_x^{\alpha} \tag{1}$$

$$Y = L_y^{(1-\beta)} K_y^{\beta} \tag{2}$$

$$Z = L_z^{(1-\gamma)} K_z^{\gamma} \tag{3}$$

where L_x, L_y and L_z represent the amounts of labor used in production of X, Y and Z, respectively, and K_x, K_y and K_z the corresponding amounts of capital. The total quantities of labor (L) and capital (K) available to the economy are assumed to be fixed.

From our assumptions above, we can obtain the following relationships:

$$L_x P_L = X P_x (1-\alpha) = \text{Cab}\,(1-\alpha) \tag{4}$$

$$K_x R_x = X P_x \alpha = \text{Cab}\,\alpha \tag{5}$$

$$\begin{aligned} L_y P_L &= Y P_y (1-\beta) \\ &= \text{Ca}\,(1-b)(1-\beta) \end{aligned} \tag{6}$$

$$K_y R_y = Y P_y \beta = \text{Ca}\,(1-b)\beta \tag{7}$$

$$L_z P_L = Z P_z (1-\gamma) = C(1-a)(1-\gamma) \tag{8}$$

$$K_z R_z = Z P_z \gamma = C(1-a)\gamma \tag{9}$$

Here P_L is wage and R_x, R_y and R_z are returns per unit of capital in X, Y and Z, respectively, and all parameters except b, namely, C, a, α, β, γ, K and L have given fixed values,[1] while b is a stochastic variable and the realized value of b differs from period to period.

Since the total revenue of industry Z does not contain the parameter b, earnings of both labor and capital of this industry will be constant over different time periods, but earnings of labor and capital in the X and Y industries depend on the realized value of b.

Clearly investing capital on production of X and Y involves risk, and on production of Z does not. We want to know how the given amounts of capital and labor available to the economy will be distributed among productions of these three goods. To anticipate our conclusions, the ratios of the distribution depend on investors' attitudes towards risk, production functions of the industries, consumers' taste and the values of the stochastic variable b and its distribution.

The Equilibrium Conditions. Suppose that the mean distribution of b is $\bar{b}$ and the standard deviation is σ_b, which of course equals $\sigma_{(1-b)}$, standard deviation of $(1-b)$. Here the correlation between random variables b and $(1-b)$ is minus one; that is, $r_{b,1-b} = -1$. We can then obtain from (5) the expected return per unit of capital from investment in industry X, and its associated risk, denoted by $\bar{R}_x$ and σ_x, respectively:

$$\bar{R}_x = \frac{Ca\alpha\bar{b}}{K_x} \tag{10}$$

$$\sigma_x = \frac{Ca\alpha\sigma_b}{K_x} \tag{11}$$

Similarly, expected earnings of per unit of capital in Y industry, $\bar{R}_y$, and its associated risk, σ_y, are given by:

$$\bar{R}_y = \frac{Ca\beta(1-\bar{b})}{K_y} \tag{12}$$

$$\sigma_y = \frac{Ca\beta\sigma_{1-b}}{K_y} = \frac{Ca\beta\sigma_b}{K_y} \tag{13}$$

Earnings of capital in Z industry, R_z, is sure return and

---

[1] Returns are measured in cash flow; that is, before deducting depreciation. If, for example, one unit of capital is \$100, and earns a net return of 10 percent in period $t+1$, $R=\$110$, not \$10. Although return is more commonly defined as net return (return after depreciation) rather than gross return, I cannot see that the different definition of return makes any difference in our analysis of investors' decisions. R defined as gross return simplifies the analysis in my model.

$$R_z = \frac{C(1-a)\gamma}{K_z} \tag{14}$$

$$\sigma_z = 0 \tag{15}$$

As seen from the equations above, both the expected returns and risks of investments in X and Y are functions of the amount of capital invested in the respective industries. (The sure return to capital in Z is a function of capital invested in Z.) Therefore, expected returns and risks in investment in X and Y are uniquely related; namely,

$$\overline{R}_x = \frac{\bar{b}}{\sigma_b}\sigma_x \tag{16}$$

and

$$\overline{R}_y = \frac{1-\bar{b}}{\sigma_b}\sigma_y \tag{17}$$

These relationships (16) and (17) are represented in an expected income-risk space in Figure 1 as two lines, denoted OT_x and OT_y for X and Y industries, respectively. For example, any point on the line OT_x, say A, represents a corresponding amount of capital investment in industry X, K_x. As the quantity of capital investment in X increases, the expected return and risk per unit of capital diminishes and this change is represented by a movement of point A towards the origin. Conversely, as the amount of capital invested in X diminishes, the corresponding return and risk increase and accordingly the point representing the new combination of the expected return and risk travels upwards along the OT_x line. The same applies to a point on the line OT_y, B, representing a corresponding amount of capital investment in industry Y. The quantity of capital invested in Z is reflected by a point C on the vertical axis or OT_z. As capital in Z increases, return per unit of capital diminishes resulting in a shift of point C towards the origin. Since the total capital available to the economy is fixed (i.e., $K_x + K_y + K_z = K$), all three points A, B and C are interrelated, and a movement of one point must be matched by a movement of at least one other point; but from the point of view of an individual investor who is in a position of the so-called "price taker," these points appear as given data.

If an investor diversifies his portfolio between sure investment (investment in Z) and risky investment (investment in either X or Y), the expected return and risk of his portfolio is represented by a point on either line AC or line BC, and the location of the point depends on the ratio of distribution of capital between C and A (or B). We name these lines the opportunity lines.

If an investor diversifies his portfolio among risky investment (investment in X and Y), his portfolio is represented by a point on a kinked opportunity line ADB, because return per unit of investment in X and Y are inversely correlated. As a proportion of investment in X in the investor's portfolio increases, a point representing $\overline{R}$ and σ of his portfolio travels on the line ADB from B to A through D.[2]

---

[2] A kinked opportunity line ADB is obtained as follows: Suppose that w proportion of the investor's capital is invested in X and the remainder $(1-w)$ in Y, the expected rate of return of his portfolio is given as follows

$$\overline{R}_{xy} = w\overline{R}_x + (1-w)\overline{R}_y \tag{18}$$

As the correlation between $\overline{R}_x$ and $\overline{R}_y$ is minus one, we can show the standard deviation of return of his portfolio as follows

$$\sigma_{xy} = \left| w\sigma_x - (1-w)\sigma_y \right| \tag{19}$$

By eliminating w from equations (18) and (19) and replacing $\overline{R}_x$, $\overline{R}_y$, σ_x and σ_y with their formulae from (10) to (13) we obtain

(1) The segment AD:

When $w(\sigma_x + \sigma_y) - \sigma_y \geq 0$ or $w \geq \dfrac{\sigma_y}{\sigma_x + \sigma_y}$

$$\overline{R}_{xy} = \left\{ \frac{\alpha b K_y - \beta(1-\bar{b})K_x}{\alpha\sigma_b K_y + \beta\sigma_b K_x} \right\} \sigma_{xy} + \frac{C\alpha\alpha\beta}{\beta K_x + \alpha K_y} \tag{20}$$

(2) The segment DB:

If an investor diversifies his portfolio among all three industries (X, Y and Z) his portfolio will be represented by a point located in either the area enclosed by the lines AC and AD or the area enclosed by the lines DB and CB on the $\bar{R}-\sigma$ space, since such a portfolio must correspond to a point on a line connecting a point on the kinked line ADB and the point C.

Before proceeding with our analysis, we introduce assumptions about investors' behavior. Suppose that there are n investors in the economy (n is a large number) and ith investor, having K_i units of investment funds

$$\left(\sum_{i=1}^{n} K_i = K\right)$$

views the outcome of his investment in a probabilistic distribution. Following the tradition of this type of economic analysis, let us assume that he is sensitive to only two parameters of this distribution; namely, the expected value and the standard deviation. Then the utility function of the investor can be represented by:

$$U = f(\bar{R}, \sigma)$$

where $\bar{R}$ indicates expected earnings of the next period (t_1) from the investment made in the current period (t_o) and σ, the standard deviation of the possible divergence of actual earnings from $\bar{R}$. (This construction of the utility function assumes that complete reorganization of investment is possible in each period for the subsequent period and, therefore, the returns of the periods beyond t_1 are independent of the investment decision of the period t_o.) The presence of either of the following two conditions justifies this mean-standard deviation approach.[3] First, the utility function of an investor is quadratic; second, an investor considers that the outcomes of an investment are distributed normally.

Let us assume that the ith investor has a quadratic utility function, such as

$$U(R) = A_iR - B_iR^2 + D_i$$

where

$$R \leq \frac{A_i}{2B_i}, \quad A_i > o \quad \text{and} \quad B_i > o$$

Thus we assume an investor is a risk-averter. The characteristics of this investor's indifference map are summarized as follows.

1) $$\frac{d\bar{R}}{d\sigma} > o$$

2) $$\frac{d^2\bar{R}}{d\sigma^2} > o$$

3) For a given σ as $\bar{R}$ increases,

$$\frac{d\bar{R}}{d\sigma} \text{ also increases.}$$

The indifference maps of all investors are now superimposed on Figure 1, where, for simplicity, only ith and jth investors' indifference maps are shown. It is clearly seen that with the initial locations of points A, C and B the most desirable portfolio from any investor's viewpoint must be located on the opportunity frontier line AC only.

This means that if in fact the choice is made between points on the line AC, the economy is not in equilibrium. All investors will want to invest in the Z and X industries, and none in the Y industry. The zero amount of capital in the Y industry is not consistent with the present

---

When $w(\sigma_x + \sigma_y) - \sigma_y < 0$ or $w < \dfrac{\sigma_y}{\sigma_x + \sigma_y}$

$$\bar{R}_{xy} = -\left\{\frac{\alpha b K_y - \beta(1-b)K_x}{\alpha\sigma_b K_y + \beta\sigma_b K_x}\right\}\sigma_{xy} + \frac{C\alpha\alpha\beta}{\beta K_x + \alpha K_y} \tag{21}$$

[3] See J. Tobin [9, pp. 63–86].

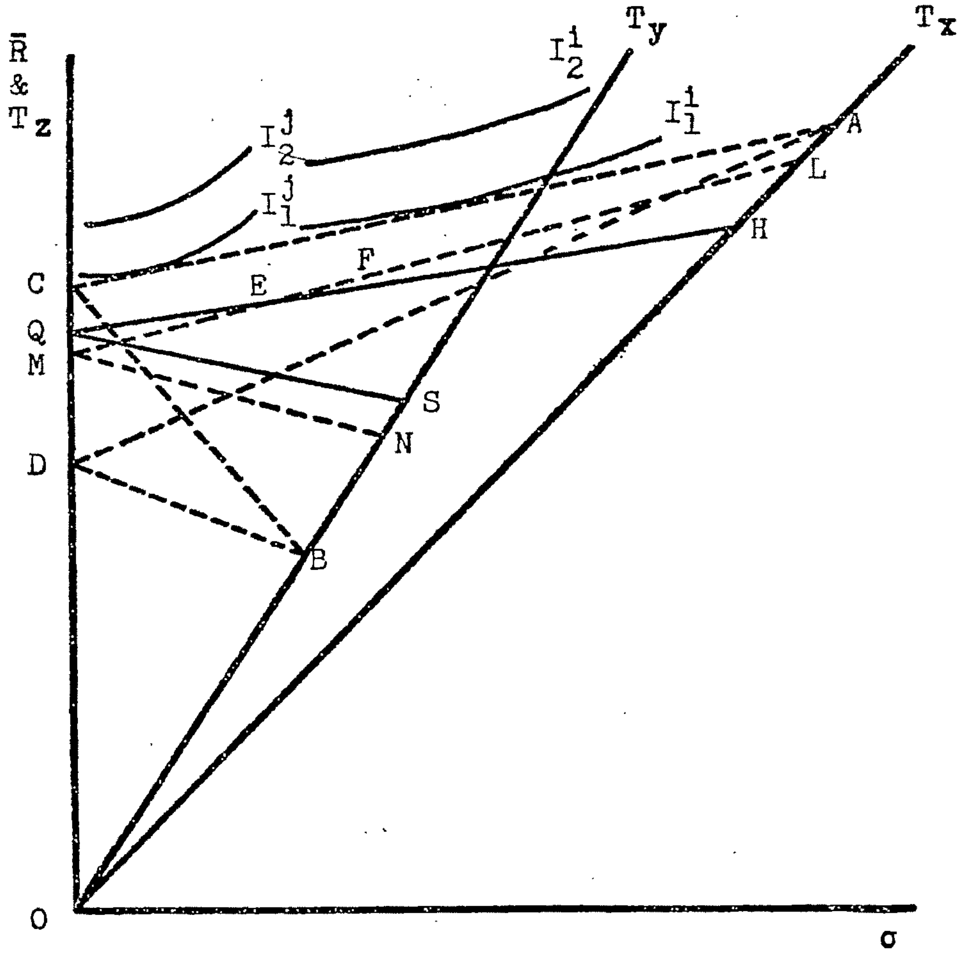


Figure 1

location of B; it implies a higher expected return per unit of capital in Y, $\bar{R}_y$, and a higher risk σ_y (see (12) and (13)). Conversely, the larger amount of capital for the X and Z industries is not consistent with the present location of A and C; it implies lower $\bar{R}_x$, σ_x and R_z (see (10), (11) and (14)).

The combined actions of all investors intending to divert their capital to the X and Z industries from the Y industry press A and C down towards the origin along the line OT_x and the $\bar{R}$ axis, respectively, and push B upward along the line OT_y. These movements of points A and B are also accompanied by an upward movement of D along the $\bar{R}$ axis. Only when C and D coincide will equilibrium investment in X, Y and Z be realized. In short, a necessary condition of equilibrium is that the opportunity lines ACB and ADB must coincide and be represented by an identical opportunity line such as LMN. Among the portfolios represented by points on LMN, all those located on the MN segment are inferior to all of those located on the LM segment. We call LM the efficient opportunity line.[4]

[4] Any portfolio represented by a point on the line LM can be obtained by diversifying investment either in X,

The efficient opportunity line LM is expressed as a function of K_z as follows.[5]

$$\overline{R} = \left\{\frac{C(1-a)\gamma + Ca\alpha\bar{b} + Ca\beta(1-\bar{b})}{Ca\sigma_b|\alpha-\beta|} - \frac{C(1-a)\gamma K}{Ca\sigma_b|\alpha-\beta|K_z}\right\}\sigma + \frac{C(1-a)\gamma}{K_z} \tag{22}$$

Here we assume that $\alpha \neq \beta$.[6]
This opportunity line LM rotates counterclockwise around a central point as capital moves from the risky industries (X and Y) to the sure industry (Z). This fact can be shown by rearranging the equation (22).[7] The coordinates of the center of rotation denoted by E are:

$$\overline{R} = \frac{C(1-a)\gamma + Ca\alpha\bar{b} + Ca\beta(1-\bar{b})}{K}$$

$$\sigma = \frac{Ca\sigma_b|\alpha-\beta|}{K}$$

Thus E represents the economy's expected return per unit of capital and the economy's risk per unit of capital. This is intuitively clear: because the expected return and the risk of the economy as a whole is determined once values of $\bar{b}$ and σ_b are given, and investors' behavior cannot influence E's coordinates. In other words, at the equilibrium situation, the weighted average of all investors' portfolios must be the same as the economy's expected return and risk, represented by point E. This is another necessary condition for the equilibrium.

In order to reach this equilibrium the economy's capital must be distributed among three industries in a specific way. Since any slope of LM represents a corresponding distribution pattern of capital among the three industries, there must be one slope which equates the weighted average of all investors' portfolios and the economy's expected return and risk.

Suppose that at a given initial slope of ML, the weighted average of all investors' intended portfolios, each of which is, of course, chosen by an investor locating a tangency point of his indifference curve to the opportunity line ML, happens to be F. The location of F (right of E) indicates that investors as a whole intend more capital in X (less capital in Z and Y) than the economy can accommodate. The excess demand for investment in X turns the opportunity line clockwise, and with the resulting change in the slope of ML all investors are forced to readjust their portfolio: the weighted average of their portfolios represented by F cannot be realized. Since all investors are risk-averters and their utility functions are quadratic, their reactions to the new opportunity line which has a smaller slope than LM produce a new weighted average point which must be closer to E than F.[8]

Y and Z or in X and Z, or in X and Y. A portfolio represented by the L can be obtained by investing in X alone and that represented by point M, by diversifying investment either in X, Y and Z or in X and Y or in Z alone.

[5] Line LM is obtained as follows. By equating R_z from (14) to $\overline{R}_{xy}$ at zero value of σ_{xy} from (20) or (21) and recalling $K = K_x + K_y + K_z$, we obtain

$$K_x = \frac{-(a\alpha\beta + (1-a)\gamma\alpha)K_z + (1-a)\gamma\alpha K}{(1-a)\gamma(\alpha-\beta)}$$

$$K_y = \frac{(a\alpha\beta + (1-a)\gamma\beta)K_z - (1-a)\gamma\beta K}{(1-a)\gamma(\alpha-\beta)}$$

Substituting for K_x, K_y in (20) or (21) the above values, we get the opportunity line LM, represented by (22).

[6] If $\alpha = \beta$ all investors will possess investment portfolios with zero risks. The Figures 1 and 2 are drawn by assuming $\alpha > \beta$. The relative position of L and N will be reversed if $\alpha < \beta$.

[7] Equation (22) can be rearranged as follows

$$\overline{R} - \frac{C(1-a)\gamma + Ca\alpha b + Ca\beta(1-b)}{K}$$
$$= \left\{\sigma - \frac{Ca\sigma_b|\alpha-\beta|}{K}\right\} \cdot \left\{\frac{C(1-a)\gamma + Ca\alpha\bar{b} + Ca\beta(1-\bar{b})}{Ca\sigma_b|\alpha-\beta|} - \frac{C(1-a)\gamma K}{Ca\sigma_b|\alpha-\beta|K_z}\right\}$$

Thus when K_z changes (i.e., $K_x + K_y$ changes) the opportunity line rotates around E.

[8] Since the investors have quadratic utility functions, the locus of the points on the ith investor's indifference curve with the same slope θ_i is expressed by

Suppose that when the slope of the opportunity line is represented by HQ (corresponding investment in Y is expressed by S on OT_y line), the weighted average of all investors' portfolio coincides with E. Then the equilibrium distribution of capital is established among three industries as indicated by three points H, Q and S.[9] This equilibrium is stable. If the opportunity line has smaller slope than HQ, the weighted average of all investors' portfolio will locate to the left of E, and the investors' intention to invest more capital in Z and Y, and less in X (than the economy can realize) will cause a counterclockwise rotation of the opportunity line.

This shows that once the production functions, consumers' taste, the amount of capital and labor are given, allocation of capital between the risky and the sure industries is determined by the shapes of the indifference curves of all investors. The larger is the proportion of capital owned by investors whose indifference curves' slopes increase rapidly as risk increases around the equilibrium opportunity line (i.e., they are more sensitive to risk), the greater is the proportion of total capital that will be invested in the sure industry Z and the risky industry Y, which has a lower expected return than X (and the smaller is the proportion of capital invested in the risk industry X which has the highest expected return of all). Conversely, the larger is the proportion of capital owned by investors whose indifference curves' slopes increase less rapidly around the equilibrium opportunity line, the smaller is the slope of the equilibrium opportunity line, the smaller is the proportion of the total capital that will be distributed to Z and Y (and the greater is the proportion of capital invested in X).

The Effects of the General Sales Tax on Risk-taking

Assume that a general sales tax is levied on sales of all three products X, Y and Z, including sales to the government, at the rate $(1-t)$, that the available quantities of labor, L, and capital, K, are not affected by the tax, and that the government spends the tax revenue in the same way as consumers would have done if there were no tax (that is, $(1-a)$ on Z, ab on X, $a(1-b)$ on Y), and distributes the resulting goods and services free of charge.[10] But the income effect on the recipients who presumably include capitalists is not taken into account in the present analysis.[11] Full employment of labor and capital is assumed before and after the tax.

When a tax of $(1-t)$ is imposed, the total gross product of the three industries X, Y and Z remains unchanged. But the after-tax return per unit of capital invested in the sure industry, Z, is reduced by the rate of the tax, $(1-t)$, and the after tax expected returns per unit of capital invested in the risky industries, X and Y, as well as their standard deviations of the outcomes (i.e., risks), are also reduced by the same proportion $(1-t)$. This means any point on HQ in Figure 2 moves downwards along a line connecting that point and the origin. For example, an investors' after-tax situation of his pre-tax portfolio, I, is represented by point I_t, and $OI_t = t(OI)$.

Thus the tax shifts the pre-tax opportunity line HQ towards the origin, keeping

---

$$\bar{R} = \frac{A_t}{2B_t} - \frac{\sigma}{\theta_t} \tag{23}$$

Therefore the opportunity line with the slope θ_t intersects the line expressed by (23) at an angle of 90° If

$$\theta_t < \theta_{t+1}, \qquad -\frac{1}{\theta_t} < -\frac{1}{\theta_{t+1}}$$

Then a point where the opportunity line of the slope of θ_t is tangent to an indifference curve is southwest of a point where the line (23) of a slope $-(1/\theta_{t+1})$ cuts the opportunity line of the slope θ_t.

[9] By our specifications of the shape of the indifference curves, the equilibrium opportunity line cannot be as steep as the OT_x line.

[10] For discussion of this assumption, see A. C. Harberger [2].

[11] For discussion of this point, see C. S. Shoup [8].

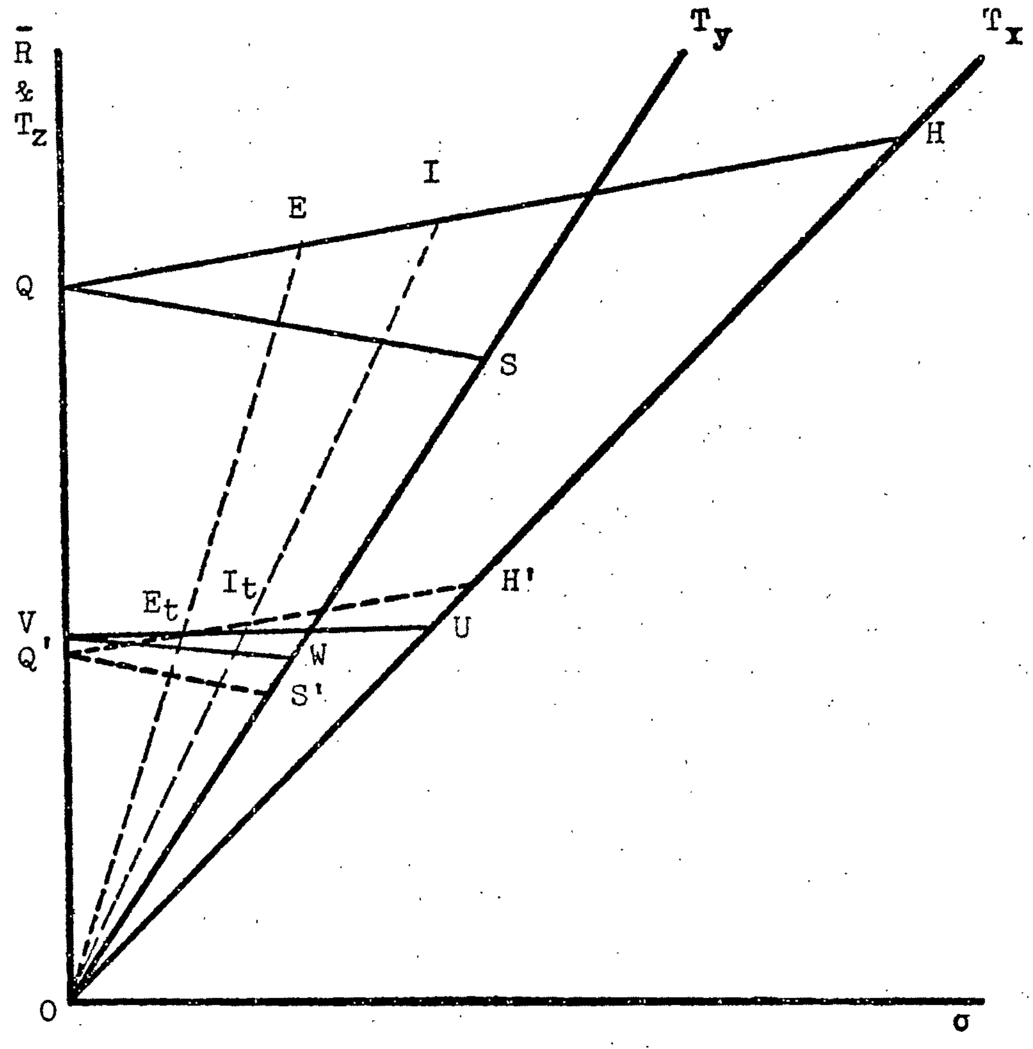


FIGURE 2

the slope unchanged. The post-tax opportunity line is denoted by $H'Q'$. The coordinates of E are also reduced by $(1-t)$ and we obtain E_t's coordinates as follows:

$$\bar{R}_t = \left\{\frac{C(1-a)\gamma + Ca\alpha\bar{b} + Ca\beta(1-\bar{b})}{K}\right\} t$$

$$\sigma_t = \left\{\frac{C a_{ab} |\alpha-\beta|}{K}\right\} t$$

Once again, the post-tax equilibrium will be reached when the weighted average of all investors' portfolios coincides with E_t. Since we assume a quadratic utility function for any investor, the slope of his indifference curve at a point such as I_t (on the post-tax opportunity line $H'Q'$) must be smaller than the slope of the indifference curve at his pre-tax equilibrium point I, for I_t is southwest of I. Then he is induced to shift his portfolio to one with higher expected return and higher risk. The same applies to all other investors. As a result the post-tax weighted average of investors' portfolios deviates from E_t and moves to the right.

Using the same argument as before we can show that the post-tax equilibrium opportunity line, such as UV, must have a smaller slope than $H'Q'$.

It is, therefore, concluded that at the post-tax equilibrium, capital investment in the sure industry (Z), K_z, must be less than the amount of capital investment in that industry at the pre-tax equilibrium. It is also observed that within the risky industries capital moves from the Y industry, the lower expected return industry, to the X industry, the higher expected return industry.

Since the vertical distance between the pre-tax opportunity line and the post-tax line becomes greater, the higher the tax rate $(1-t)$ is, the higher the rate of the sales tax is, the more capital is shifted from the sure industry Z to the risky industry X, and within the risky industries more capital moves from the Y industry to the X industry.

REFERENCES

1. E. D. DOMAR AND R. A. MUSGRAVE, "Proportional Income Taxation and Risk-Taking," *Q.J.E.*, May, 1944, pp. 387–422.
2. A. C. HARBERGER, "The Incidence of the Corporation Income Tax," *J.P.E.*, June, 1962, pp. 40–65.
3. H. M. MARKOWITZ, "Portfolio Selection," *J. of Fin.*, Mar., 1962, pp. 77–91.
4. J. MOSSIN, "Taxation and Risk-Taking—An Expected Utility Approach," *Economica*, Feb., 1968, pp. 74–82.
5. R. G. PENNER, "A Note on Portfolio Selection and Taxation," *Rev. of Econ. Studies*, Jan., 1964, pp. 83–86.
6. M. K. RICHTER, "Cardinal Utility, Portfolio Selection and Taxation," *Rev. of Econ. Studies*, No. 3, 1960, pp. 79–93.
7. W. F. SHARPE, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *J. of Fin.*, Sept., 1964, pp. 425–42.
8. C. S. SHOUP, *Public Finance* (Aldine Pub. Co., forthcoming).
9. J. TOBIN, "Liquidity Preference as Behaviour towards Risk," *Rev. of Econ. Studies*, Feb., 1958, pp. 65–86.

DISCUSSION

DAVID CASS: It's Saturday night after a particularly trying week. I get the urge for a cold beer, hop in the car, and run down to the corner drugstore for a sixpack.

Now it's early morning during a hot, muggy summer spell. The air conditioner has just shorted, starting a small but easily controlled fire. My wife—still shaking—is bugging me about fire protection. I protest that there's a firehouse just down the street as well as a hydrant right next door (but end up promising her that I will go out first thing next morning and pick up some home extinguishers and also—against my better judgment—talk to an insurance agent about increasing our coverage).

What do these situations have in common? Zeckhauser would have us believe it is the types of commodities, or, more precisely, the uncertain demand for the commodities involved—cold beer and fire-fighting services. Contrary to his implicit assertion, the necessity of "decisions on the provision of a good before preferences become known" does not really distinguish cold beer on a whim from a pumper on beck and call. What does is that uncertainty plays such a minor role in allocation questions involving most commodities,[1] such a major role in those involving large but unlikely losses in a game against a malevolent but impersonal nature. The latter is surely what Zeckhauser is primarily concerned with.

Well, what is different about efficient (i.e., Pareto optimal) allocation under uncertainty? Following the pioneering lead of Arrow and others, I assert that it is nothing more or less than the fact that efficient allocation requires the existence, in sufficient variety, of markets for contingent claims. And while this is suggested both in his extensive example of a community's provision against loss by fire and again later in his discussion of "Insurance and Markets for Contingent Claims," it seems to me that Zeckhauser simply doesn't bring out this central point, along with some of its more obvious implications, with sufficient force and clarity. (Indeed, several parts of his discussion appear to me pretty much beside the point; see, in particular, (1) "Utilities of Anticipation versus Utilities of Use," a distinction both reminiscent of last century's rather useless debates on the meaning of demand as well as superfluous under the expected utility hypotheses or its kin, and (2) "Collective Provision of PIP Goods," a misleading discussion of the problems associated with indivisibility and increasing returns—problems no more intrinsic to the analysis of allocation under uncertainty than to that of any other set of allocation questions.[2]) I will take this opportunity to expand a bit on his example, and attempt to make up for this—in my view—rather serious lack of focus.

The example, as originally formulated, does illustrate the fact that a market for contingent claims may be necessary to insure efficient allocation under uncertainty, in this instance taking the form of a market in which $\$x$ now trades for \$10,000 if a particular house burns. Notice, however, that services of the "super fire engine" could also be traded in a market for contingent claims. That any such market would be inoperative under the particular circumstances Zeckhauser postulates doesn't alter the fact that its operation might be necessary to insure efficiency. Suppose, for example, that the one hundred individuals are all risk-lovers rather than risk-averters, that each individual's demand price for \$10,000 if his house burns down is \$89, and finally that the rental for the "super fire engine" is \$9,000 rather than \$12,000. Clearly under these circumstances the insurance market mentioned above will be inoperative, while what Zeckhauser refers to as "conventional efficiency dictates" will rule out any collective decision to rent the fire engine. Yet just as clearly the individual whose house will burn otherwise will prefer an arrangement whereby he alone pays not more than $10,000 for the services of the engine if his house catches fire. But availability of this opportunity is nothing more than the existence of a particular market for contingent claims, one in which $$x$ if a particular house catches fire trades for services of the engine if that particular house catches fire.

---

[1] This is not a denial that there are any interesting issues raised by the fact that I don't ordinarily order my beer six months in advance. Only that they're of a different order than those raised by the fact that I have no idea when my air conditioner might start an electrical fire.

[2] Fires in an urban area do involve externalities while fire-fighting technology does exhibit indivisibilities and increasing returns; however, as Zechhauser correctly implies by his device of a "super fire engine," these are not fundamental to the problems raised by the uncertainty of the incidence of fire from the individual's or society's viewpoint.

This permutation of the example, besides illustrating my basic point—I should emphasize especially that both commodities traded are contingent claims—also has implicit in it the fact that our usual notions of competitive, maximizing behavior require modification in the context of markets for contingent claims. (The original example also involves this consideration.) For, given such a market, how could any private firm provide the services of the engine (or for that matter insurance) at nonnegative profit except by ending up with all the individuals as customers? And, an even prior question, how could a private firm realistically calculate the profitability of transacting in such a market except with knowledge of both price and quantity? While it's clear that one can think of ways of getting around this difficulty which make sense in this specific example (e.g., an arrangement whereby several firms bid, each on the assumption that it will end up with either zero or 100 customers), this aspect of competitive maximizing behavior in markets for contingent claims seems both unavoidability related to questions of efficient allocation under uncertainty and very much in need of a good deal of further analysis.

One can easily think of other permutations of the example which bring out yet further difficulties associated with questions of the efficiency of allocation under uncertainty. One obvious difficulty, already suggested, is that the number of markets (operative and inoperative) necessary for efficiency may be very large indeed. This point can be highlighted in the context of this example simply by assuming that all individuals are not identical (e.g., fifty-one individuals are risk-lovers with average demand price for $10,000 if one's house burns of $90, forty-nine risk-averters with average demand price for $10,000 if one's house burns of $111 and the rental of the "super fire engine" is greater than $10,029). What determines whether such markets exist? What are the costs associated with the availability and use of these markets? And so on.

Another related problem is that the contingent claims involved may be very intricate—a point aptly illustrated by considering the claims required if both each individual's loss from fire and all individuals' loss from fire are uncertain (e.g., Zeckhauser's example in the section "The Community Insurance Decision When the Size of the Loss Is Variable"). What determines whether a requisite claim is invented? How well do less complicated proxies perform? And so on.

Obviously I can only scratch the surface of this area here. But hopefully my discussion serves to emphasize the importance as well as suggests some of the difficulties of the viewpoint that the critical characteristic—and I repeat myself—of efficient allocation under uncertainty is the existence in sufficient variety of markets for contingent claims.

WILLIAM H. OAKLAND: My remarks will be confined to Dr. Shibata's paper on taxation and risk-taking.

Since virtually all existing theory of portfolio selection under uncertainty has been partial equilibrium in character, this paper, which examines the interaction between investor's decisions and the allocation of real capital, represents an important and refreshing departure from tradition. Viewed from this perspective, theorems concerning the effects of taxation upon risk-taking, while interesting, are incidental rather than of central interest. What is important are the properties of the model within which the analysis takes place. It is from this point of view that I have prepared my remarks. However, it is interesting to note that in the simple economy dealt with in this paper, a general income tax is equivalent to a general sales tax. Since there is no depreciation or negative quasi-rents, the usual distinction which rests on loss offsets does not matter. Thus, this paper can be viewed as an extension of the theory of the effects of income taxation upon risk-taking.

As a model of general equilibrium, Professor Shibata's contribution is somewhat incomplete. Several key relationships are left unspecified. For example, it is asserted that the level of money income is constant over time. Without the explicit introduction of money into the model, it is not clear why this should be the case. Furthermore, some allocations of capital are more efficient than others. That is, with a constant price level, the level of income will be affected by a reshuffling of capital among industries. By ignoring such income-effect feedbacks, the analysis is enormously simplified. But the cost of such a simplification is difficult to identify.

Also of significance is the lack of a market for capital goods. Firms, as producing units, are assumed to employ whatever capital is offered them and at any price. This follows from the fact that owners of capital, called "investors," allocate real capital to the various sectors based upon expected returns and risks. Firms, in turn, are assumed to pay out whatever quasi-rents are earned. As a consequence, there is absolutely no reason to distinguish between firms and investors.

By itself, this lack of distinction creates no problem since capital and labor are perfectly divisible and production takes place under constant returns to scale. Each investor can be viewed as a multiproduct firm. However, this treatment would appear to preclude the exchange of capital among investors. In other words, debt finance is not an available alternative to a firm. If the quasi-rent on capital invested in the "sure" industry were invariant, it could serve as contractual debt in the model. However, the return to investment in the sure industry depends upon the level of investment in that industry, and the latter is not known with certainty.

This leads me to a third point: the formation of expectations in the model. Given the Cobb-Douglas production, the realized quasi-rent per unit of capital in industry s is given by

$$R_s = v_s C/K_s \tag{1}$$

where v_s is the share of national income expended upon good s. For fixed value of C and K_s, the distribution of R_s will be proportional to the distribution of v_s. In equilibrium, K_s is constant so that the latter obtains. It is reasonable to assume that investors, by empirical observation on quasi-rents, could learn the distribution of R_s. In situations of disequilibrium, however, the distribution of R_s is conditional upon the value of K_s. Since the latter does not vary in a systematic way it is difficult to imagine that investors would be able to determine the distribution of R_s by empirical observation alone. Dr. Shibata implicitly assumes that investors know the right-hand side of (1) and that they have static expectations with respect to K_s. Only when K_s remains fixed, therefore, will investors correctly anticipate the distribution of R_s.

This procedure appears to be objectionable on several grounds. First, if investors do know the exact form of (1), it seems arbitrary to assume that they form sophisticated expectations about v_s and naïve assumptions about K_s. Second, but more important, (1) is a reduced form equation not likely to be known by the individual investor. If he forms expectations on R_s, it is likely to be based upon empirical observation. However, in disequilibrium he will not be successful in deducing a stable distribution of R_s. Each investor will form his own subjective estimates of the true distribution. In particular, each owner-investor will recognize that the quasi-rents earned in industry s depend only upon the price of output and the wage rate which prevail in the future. Specifically,

$$R_s = P_s^{1/\rho} P_l^{(\rho-1)/\rho} \rho (1-\rho)^{-(\rho+1)/\rho} \tag{2}$$

where ρ is the elasticity of output with respect to capital. Furthermore, it can be deduced that P has the distribution

$$P_l = H + b(\beta - \alpha) C/L \tag{3}$$

where H is a constant. Thus the mean and variance of the wage rate are constant over time. It is reasonable to suppose that investors will know the exact distribution of the wage rate. All that remains is the specification of price expectations. Since realized price is given by

$$(4) \qquad P_s = v_s C/S$$

where S is total industry output, it might appear that investors could use this relation to form their price expectation. However, the same objection applies to this line of reasoning as to the use of (1)—S does not remain constant. It is significant to note, however, that even if investors had access to (4), the combination of (2), (3), and (4) would not yield (1). This casts the use of (1) further into doubt.

Returning to the argument, each investor will form his own price expectations, say based upon the average value of S. In general such expections will not give rise to the solution generated in the paper.

My next comment has to do with the use of the nominal quasi-rent in the utility function of investors. This procedure would be acceptable if the aggregate price level were constant. However, in order to insure a constant level of money income, the aggregate price level must vary. This is due to the fact that as the sectoral allocation of the capital stock is changed, the level of real income will change. Real income will be maximized if capital is allocated such that the realized quasi-rent is the same in all industries. It is interesting to note that on the average the solution offered in the paper does not satisfy this criterion. Since the aggregate price level will fluctuate, it appears desirable to assume that investors formulate their utility in terms of the real quasi-rent rather than the nominal quasi-rent.

The main thrust of my comments is that the proposed model is satisfactory only if it is assumed that the economy is in a constant state of equilibrium. Nothing can be said concerning the adjustment to equilibrium or the stability of equilibrium with the present framework. Nevertheless, this paper provides a useful first step in the development of a general model of portfolio behavior under uncertainty.

Mark V. Pauly: Economists have, within the last decade or so, made considerable progress in extending the application of economic theory to analysis of the behavior of governmental bodies. Mr. Devine's paper is intended as a contribution to that literature. It looks at a common problem which has not heretofore been subject to much theoretical analysis—"shortages" of certain kinds of local government employees—and attempts to explain it.

This is a welcome endeavor. But I have some doubts about the way in which Devine approaches the problem. The model he uses was designed to explain the existence of equilibrium shortages and expenditure on recruiting of employees by a profit-maximizing nondiscriminating monopsonist. I wonder if this is the right model to use when analyzing a not-for-(explicit)-profit local government.

The reasons why the profit-maximizing monopsonist experiences shortages are presumably (1) because he cannot distinguish employees in such a way as to discriminate among them and (2) because his profits would be reduced if he raised recruiting expenditure and wages. But does a local government exhibit either of these characteristics? It need not be explicitly concerned with maximizing profits or minimizing costs. Its demand curve for labor is not a value-of-marginal-product curve. Rather, its demand curve for labor is determined through the political process, both indirectly as a derived demand for the voter expressed demand for government services and directly through "political" bargaining between governments and employee groups. So I see no obvious reason why a government might not offer a wage equal to or even greater than that needed to eliminate shortages for types of workers in which it is a major purchaser.

It is possible, however, to outline a model of bureaucratic decision making for which the monopsonist framework is appropriate. Suppose we assume that heads of departments are given a budget each year, and that they then desire to maximize the number of persons under them which can be hired for that budget. They will set wages at the monopsony level but will be willing to hire more workers at that level.

Second, one might ask why a local government may not discriminate. In fact, higher pay for those with more years of service, which occurs in some municipal agencies, might be an example of discrimination that does occur. It is true, as Devine notes, that there are some constraints on discrimination, but these are not the technical ones which may make the model of the monopsonist relevant for some profit-maximizing firms but rather they are political constraints, arising within the same process that determines budgets and wage rates. As such, one might have hoped that they would have been included in the model rather than being exogenous to it. I shall return to this point when discussing Devine's empirical examples.

A third theoretical objection is concerned with the meaning of the term "vacancy." The existence of an official vacancy may not really reflect a desire to hire more employees. If, as Devine's model suggests, there can be permanent or equilibrium vacancies—vacancies which will never be filled—then the number of vacancies is really meaningless in terms of behavior and can be manipulated at will. A vacancy need not reflect the willingness to hire additional employees if it is known that no applicants will appear. If bureaucrats desire to maximize their budgets or the number of employees under them, one way to get a bigger budget is to have a lot of "vacancies." Since the marginal cost of creating an unfilled vacancy is zero, we might expect bureaus to try to appear permanently understaffed. The difficulty with this answer is that it would appear to suggest shortages in all job classes. My only explanation of observed behavior is that in some classes where the service rendered is nonspecific (e.g., teaching and police), it may be easier to fabricate shortages.

Finally, we ought not to ignore the possibility that shortages may simply be the result of a lag in wages behind demand and may only be a disequilibrium phenomenon.

Devine illustrates (but does not test) his model with three examples of shortages: registered nurses, police, and teachers. He first provides some tables of jobs for which shortages and queues exist, but their relevance is not clear. For the jobs of meter readers, park service attendant, and refuse collection foremen there are queues, although Devine's theory would, if generally applicable, suggest shortages. It would also suggest no shortages for "competitive" jobs like auditor and clerk typist, apparently contrary to fact.

The shortage of registered nurses exists only at hospitals where the city and county are parties to an area-wide wage-fixing arrangement. There is no shortage of nurses for schools or for public health positions, where governments are also in a monopsony position. Presumably there is no shortage because wages are better in these positions, but why are they better? Why does the government not behave as a monopsonist here? Devine's theory offers no explanation. Donald Yett suggests that it is because of the cost of adjusting the wage structure and because public health nurses are hired through civil service. If this is true, it suggests two additional elements—the cost of making alterations in the published wage structure and the use of civil service categories and procedures—that ought to be included in a complete theory of wage determination by local governments. The existence of civil service procedures raises some other interesting points. Can one predict which jobs will be put under civil service? Will they tend to be those in which the governmental unit has a monopsony or not? Devine's theory does not explain this.

Apparently the shortage of policemen results from a forced parity between wages of policemen and firemen, and there are queues of firemen. There may be a political reason for this tie, but Devine does not explain it. Why are governments bound by such constraints on the wage structure? And why is the wage of firemen tied to that of policemen, while the wage of garbagemen, for example, is not? A general theory of shortages ought to provide some explanation of this sort of thing.

As an alternative explanation, it may be that governments would dislike the long and politically explosive queue for firemen that would occur if fire-police wages were raised to a level sufficient to eliminate the shortage of policemen. It is interesting that the three examples of shortages Devine gives all would involve some sort of extensive queuing—by nurses for nonhospital positions, by firemen, or by teachers for jobs in suburban schools—if shortages were eliminated. Given the constraints on discrimination, the only way to eliminate a shortage is to create a queue.

The shortage of teachers is also apparently the result of a failure to discriminate in paying more for posts at ghetto schools. Again, the failure to discriminate would appear to arise from political rather than technical reasons. Finally, in this latter case, the shortages usually are not lack of applicants at wages offered, but lack of qualified teachers. Since the qualifications required for primary and secondary school teachers are, in many states, probably irrelevant to job performance anyway, it is not clear that this represents a genuine shortage. "Qualifications" may be only a political method of restricting supply.

In sum, it does not appear to me that Devine has either spelled out his theory as it applied to local governments very clearly, or that he has tested it and found it confirmed. *Ad hoc* "political" reasons and not his theory seem to explain shortages he does find. I do not mean to suggest that theory is not worth testing. It would clearly be useful to know, for example, if there was a relationship between the vacancy rate (or some other measure of shortage) and the fraction of a job class hired by the city or county. And it would be interesting to know how the equilibrium shortage varied over time as demand and supply conditions varied. I hope Devine will provide us with such tests in the future.

AMERICAN ECONOMIC ASSOCIATION

PROCEEDINGS
OF THE
EIGHTY-FIRST ANNUAL MEETING

CHICAGO, ILLINOIS

DECEMBER 28-30, 1968

PROCEEDINGS OF THE AMERICAN ECONOMIC ASSOCIATION

ANNUAL BUSINESS MEETING, DECEMBER 29, 1968
PICK-CONGRESS HOTEL, CHICAGO, ILLINOIS

The Eighty-first Annual Business Meeting of the American Economic Association, Chaired by President K. E. Boulding, was called to order in the Great Hall of the Pick-Congress Hotel, Chicago, Illinois, at 9:45 p.m. In his introductory remarks, the President reviewed the circumstances leading to the decision by the Executive Committee not to move the 1968 annual meeting from Chicago. In the discussion which followed, some members of the audience expressed their approval of the decision while others were highly critical. In answer to those who argued that by meeting in Chicago the Association was supporting the position taken by the mayor and other officials of the city of Chicago at the time of the Democratic Convention, the President read the following statement, adopted unanimously by the members of the A.E.A. Executive Committee present at a special meeting on the afternoon of December 29, 1968:

> A significant number of the membership of the Association have communicated to the Executive Committee their dissatisfaction with the Executive Committee's decision not to change the site of the 1968 annual meeting from Chicago to some other city. Another significant number have indicated their support of the Executive Committee's decision. Those members protesting the Executive Committee's action base their protest on the belief that holding the meeting in Chicago would constitute an endorsement of the actions of Chicago officials and police during the Democratic national convention last summer.
>
> The Executive Committee respects the views of these members and in no way wishes to abridge their freedom to express their views individually or collectively. The Executive Committee wishes to make it clear that the decision to hold the meeting in Chicago in no way constitutes an endorsement, explicit or implicit, of last summer's actions by the city of Chicago, its officials, or the demonstrators.
>
> The American Economic Association was founded as a professional society, and by constitution, tradition, and policy does not, as a society, take a position on political issues. Indeed its refusal to do so is a precondition for its individual members to express and to work for their own political and social beliefs.

After further discussion, it was VOTED to request the Executive Committee to poll members of the Association by mail regarding their preferences in respect to: (1) whether or not to change the location of the 1972 annual meeting, presently scheduled for Chicago and (2) whether or not to shift the time of future annual meetings of the Association from the period between Christmas and New Year's to late spring, early fall, or to an arrangement whereby the meeting would be held one year between Christmas and New Year's and the following year during August.

It was VOTED to approve the minutes of the Business Meeting of December 29, 1967, and to ratify the minutes of the Executive Committee meetings held on December 27, 1967, and March 8–9, 1968, including reports of officers and committees of the Association and the following restatement of the bylaws as recommended by the Executive Committee at its meeting of December 27, 1967:

> *Article III, Section 3:* The Association shall have the following officers who shall be appointed by the Executive Committee: a Secretary, a Treasurer, a Managing Editor of the *American Economic Review,* a Managing Editor of the *Journal of Economic Literature,* and a Counsel. The terms of office of each of these officers shall be three calendar years.
>
> *Article III, Section 4:* The Executive Committee shall consist of the President, the President-elect, two Vice-Presidents, the Secretary, the Treasurer, the two Managing Editors, the two ex-Presidents who have last held office, and six elected members, provided the Secretary, the Treasurer, and the two Managing Editors shall not be entitled to vote in the Executive Committee's meetings.

The Report of the Secretary was presented by H. F. Williamson. The schedule of times and places for future annual meetings through 1976 was outlined. He noted, however, that the meetings scheduled for Chicago in 1972 and 1975 were tentative, depending on further consideration by the Executive Committee. Attention was called to the growth in the Association's total number of members and subscribers from 23,305 as of December 31, 1967, to 24,713 for the same date in 1968, or a net increase of 1,408.

In addition, the Secretary reported on the following activities of the Association:

1. *National Register of Scientific and Technical Personnel.* The Association participated in the 1968 Register of Scientific Personnel. Over 22,000 questionnaires were mailed to members of the profession. Of this number,

nearly 14,000 were returned. Up-to-date information on the relative status of the profession for 1968 will be made available by the National Register during 1969.

2. *Foreign Graduate Student Screening Project.* While the project, now in its fourth year, will continue to handle requests from economics departments for evaluations of foreign students, the present plan is to discontinue the service at the end of the academic year 1968–69. Plans are well under way, however, to publish a revised *Guide to Graduate Study in Economics* in the United States early in 1969.

3. *Handbook.* Plans are well advanced for the publication of a new edition of a "who's who" type of *Handbook* late in 1969.

The Treasurer indicated that because of the shift in the Association's accounting period from a fiscal to an annual basis, financial information on operations for 1968 was not available. According to the Auditor's Report covering operations during the thirteen-month period ending December 31, 1967, however, income over expenditures amounted to $34,222, while the unappropriated surplus at the end of the period was $433,081.

In the absence of J. G. Gurley, Editor, George H. Borts, Editor-elect of the *American Economic Review,* reviewed the operations of the Editorial Office during 1968. He also summarized the future publication plans for both the *American Economic Review* and the *Journal of Economic Literature.*

The Secretary then presented the report of the Committee on Elections and the certification of the election of new officers for the year 1969 as follows:

> In accordance with the bylaws on election procedure, I hereby certify the results of the recent balloting and present the reports of the Nominating Committee and the Committee on Elections.
>
> The Nominating Committee, consisting of Joseph J. Spengler, Chairman, John M. Letiche, Ruth P. Mack, Robert V. Roosa, Phillip Saunders, and Robert A. Solo, presented to the Secretary the list of nominees for the respective offices:
>
> *For President-elect*
> Wassily Leontief
>
> *For Vice-Presidents*
> Moses Abramovitz
> Joe S. Bain
> Abram Bergson
> Irving B. Kravis
>
> *For Executive Committee*
> Mary Jean Bowman
> Roland N. McKean
> Harvey S. Perloff
> Charles L. Schultze
>
> The Committee on Elections, consisting of Robert Eisner, Chairman, Walter D. Fisher, and Harold F. Williamson, prepared biographical sketches of the candidates and ballots were distributed shortly after the middle of September. The canvass of the ballots was made on November 11, 1968, and the results were filed with the Secretary.
>
> From the report of the Committee on Elections, I have the following information:

Number of envelopes without name for identification	205
Number of envelopes received too late	33
Number of defective ballots	—
Number of legal ballots	5,395
Number of returns from the mail ballot	5,633

> On the basis of the canvass of the votes cast, I certify that the following persons have been duly elected to the respective offices:
>
> *President-elect* (for a term of one year)
> Wassily Leontief
>
> *Vice-Presidents* (for a term of one year)
> Moses Abramovitz
> Joe S. Bain
>
> *Members of the Executive Committee* (for a term of three years)
> Mary Jean Bowman
> Charles L. Schultze

Following the report on elections, President W. F. Fellner took over the chair.

In response to suggestions from the floor, the President agreed to consider appointing a committee to analyze the operation of the job market for economists.

The President introduced President-elect Wassily Leontief, who outlined his plans for the program for the 1969 annual meeting.

In response to suggestions from the floor, Professor Leontief agreed to inquire into the feasibility of making papers to be presented on the program available to members prior to the annual meeting.

The Secretary read the following statement, which was adopted:

> John G. Gurley, as Managing Editor of the *American Economic Review* from 1963 to 1968, gave distinguished service to the American Economic Association in a period marked by both rapid expansion and by crisis. The volumes which he edited are a testimonial more eloquent than any resolution. Nevertheless, this meeting wishes to express its appreciation, on behalf of all the members of the Association, for the great service which John G. Gurley has rendered.

The Secretary then read the following statement, which was adopted:

> On behalf of all members of the economic profession, this meeting wishes to express its deep appreciation to Arthur Smithies for his pioneering service as Editor of the *Journal of Economic Abstracts* during the period 1963–68.

On behalf of the Executive Committee, the Secretary presented the following resolution, which was adopted:

> In view of the extraordinary difficulties associated with planning the 1968 annual meeting, this meeting wishes to record its special thanks to William Hocter, Chairman, and the members of the Committee on Local Arrangements for their efficient management of the meeting.

The Secretary presented the following resolution, which was adopted:

> Be it resolved that this meeting commend W. J. Fellner and members of his committee who in the face of unusual difficulties were nevertheless able to prepare and implement such an excellent and varied program for the 1968 annual meeting.

The meeting adjourned at 10:50 p.m.

HAROLD F. WILLIAMSON, *Secretary*

MINUTES OF THE EXECUTIVE COMMITTEE MEETINGS

1. Minutes of the spring meeting held in New York City, March 8–9, 1968:

The *first meeting* of the *1968 Executive Committee* was called to order at 9:35 a.m. in the New York Hilton Hotel, New York, New York, March 8, 1968. The following were present: K. E. Boulding, presiding, J. M. Buchanan, Otto Eckstein, W. J. Fellner, Milton Friedman, J. G. Gurley, W. W. Heller, Fritz Machlup, Franco Modigliani, Anthony Scott, R. M. Solow, and H. F. Williamson. Absent were: K. J. Arrow, L. R. Klein, and Arthur Smithies. Present as members of the Nominating Committee were: J. J. Spengler, Chairman, J. M. Letiche, R. P. Mack, R. V. Roosa, Phillip Saunders, and Robert Solo. Present as guests were: I. B. Kravis, Mark Perlman, and Richard Ruggles.

1. *President's Remarks* (K. E. Boulding). President Boulding outlined the items on the agenda to be covered at the meetings.

2. *Minutes.* The minutes of the December 27, 1967, Executive Committee meeting were approved.

3. *Report of the Secretary* (H. F. Williamson). A preliminary report indicated a total net registration at the Washington, D.C., meeting of approximately 10,000. It was indicated that plans were well under way for the 1968 meeting in Chicago, with William J. Hocter, of the Chicago Federal Reserve Bank, as Chairman of the Local Arrangements Committee. Following a discussion, the Secretary was instructed to reconsider the plans to hold the 1971 annual meeting in New Orleans and to investigate the possibility of meeting again in San Francisco in 1974.

Papers and Proceedings. It was announced that the plan was to print 26,500 copies of the 1968 *Papers and Proceedings* and the total number of manuscript pages would be approximately 1,255.

Clearinghouse Project. The Secretary noted that arrangements had been made to transfer the operation of the Clearinghouse Roster to the Overseas Educational Service starting in June, 1968. The Overseas Educational Service, with headquarters in New York City, is particularly well qualified in this area, having for a number of years maintained similar rosters in cooperation with the American Library Association, the American Association of Collegiate Schools of Business, and the American Society for Engineering Education.

N.S.F. Register of Scientific Personnel. It was announced that the 1968 National Register questionnaires would be mailed to members of the economics profession within the next few weeks.

Foreign Graduate Student Screening Project. The Secretary announced that an arrangement had been made with Wyn Owen, Director of the Economics Institute, to have the Advisory Committee of the Institute review the Foreign Student Screening Project.

4. *Report of the Treasurer* (H. F. Williamson). It was noted that because of the new method of reporting the income the auditors could only submit a tentative and preliminary draft of the annual report for the Association as of March 7, 1968. For this reason, the final report of the Treasurer covering 1967 will be made at the December, 1968, meeting of the Executive Committee. In view of the difficulty in completing the auditor's report before the end of March, it was decided that serious consideration should be given to the possibility of holding the spring Executive Committee meetings in April.

5. *Finance Committee* (H. F. Williamson). Following a discussion, it was VOTED to reappoint Wells Farnham Chairman and Corliss Anderson to the Finance Committee for 1968 and to add Milton Friedman as a member. The Secretary was instructed to express the thanks of the Executive Committee to the members of the Finance Committee for their services during 1967.

6. *Report of the Managing Editor, A.E.R.* (J. G. Gurley). It was VOTED to approve the nominations to the Editorial Board of the *American Economic Review* as presented by the Editor. The Secretary noted that according to the George Banta Company there would be a significant reduction in composition and paper costs if the page size of the *Review* were to be changed to 7 × 10 inches, with two columns. Following a discussion, it was decided to recommend this change to the new editor of the *Review.*

7. *Report of the Managing Editor, J.E.L.* (Mark Perlman). The Editor made the following recommendations in regard to the *Journal of Economic Literature:* (*a*) the *Journal* would appear quarterly, beginning with the issue of March, 1969; (*b*) the page size of the *Journal* will be 7 × 10 inches and the cover of white stock with a two-color ink design; (*c*) each issue will consist of four parts: (1) a lengthy commissioned and signed article on recent literature developments in a subfield of economics and one or two signed interpretive articles giving the author's views on ideas of current interest and book notes and book reviews, (2) approximately 80 conventional length book reviews plus a larger number of book notes ranging in length from 50 to 100 words, identifying the major theme of each book, (3) abstracts of approximately 250 articles identical with the substantive material presently appearing in the *J.E.A.*, (4) a list of all articles on economics appearing in the journals currently covered in the *J.E.A.* but arranged by subject according to a four-digit classification scheme; (*d*) that the managing editor be authorized to recommend individuals to be appointed by the Executive Committee to a six-man Editorial Board; and (*e*) the Executive Committee authorize for editorial preparation expenditure (including one commissioned review article of a subfield) of $12,750 during 1968 and $41,200 for the operation of the *J.E.L.* office for 1969.

Following a discussion it was VOTED to accept the report and recommendations of the Editor.

It was also VOTED that the subscription rate to the *J.E.L.*, starting January 1, 1969, should be $6.00 a year.

8. *Report on the Study of the Structure of the Profession* (N. A. Tolles). The Secretary read a communication from Professor Tolles, indicating that his Report on the Study of the Structure of the Profession would soon be ready for publication. It was VOTED, subject to approval by the committee appointed to advise Professor Tolles, to publish the report on the profession as a Supplement to the September, 1968, issue of the *American Economic Review.*

9. *Reports of Standing Committees.*

Committee on Research and Publications (I. B. Kravis). The Chairman outlined the following matters that were considered at a meeting of the Committee in New York City on February 28, 1968:

a) *Price of the Volumes in the A.E.A. "Readings Series."* The feasibility of reducing the price of future volumes in the "Readings Series" was discussed with Irvin Grimes, President of Richard D. Irwin, Inc. Among the possibilities considered were: (1) subsidization by the A.E.A., (2) foregoing by the A.E.A. of its share of the profits from publication, (3) increasing the size of the initial publication as a way of reducing unit costs, (4) the use of double-column printing to cut composition and paper costs, and (5) a greater use of paper covers instead of hard covers. It was agreed that Mr. Grimes would again consult with the Committee after exploring the foregoing possibilities.

b) *The Publication of Volumes in the "Readings Series."* The status of planned volumes in the series is as follows: International trade (Caves and Harry Johnson), in press; welfare economics (Arrow and Scitovsky), publisher expects to receive MS momentarily; labor economics (Bowen and Rees), final selection will be made in summer of 1968; agricultural economics (Fox and Gale Johnson), final selections should be made in a few months; and growth economics (Solow and Lewis), no progress has been reported on this volume and the Chairman is to inquire about the editors' intentions.

c) *Foreign Surveys.* The list of authors and countries covered for the surveys of economic policies, edited by Harry Johnson, is as follows: W. M. Corden, Australia; David W. Slater, Canada; Bronco Horvat, Yugoslavia; Sergio de Castro, Chile; Joan Knox, South Africa; Jagdish Bhagwati, India; Leopoldo Solis M., Mexico; and Assar Lindbeck, Sweden. It was agreed that it would be desirable to have a survey of Israel by Naday Halevi, provided that finances are available.

d) *The Publication of Volumes in the "Translation Series."* The projected schedule for the completed translations of volumes in this series is as follows: Pareto, *Manuel* (Dehem), translation complete except introduction and mathematical appendices; Dahmen, *Svensk Industriell Foretagarverkamhet* (Leijonhufvud), April, 1968; Tugan-Baranovsky, *The Russian Factory* (Levines, Grossman), MS expected in summer, 1968; French economic history (Cameron), spring, 1969; Galiano, *Della Moneta* (Toscano), no completion date; Japanese economic history (Crawcour), spring, 1969; Kula, *Problemy i Metody* (Szreter), spring, 1969; and Chinese writings on money (Hartwell), fall, 1969.

e) *Additional Volumes for the "Translation Series."* It was recommended that the Association approve the idea of continuing support from its own resources for the translation of foreign works of special importance and that serious consideration be given in the near future to the possibility of translating a third volume of Pareto's works.

Following a discussion, it was VOTED to accept the report of the Committee on Research and Publications, including the recommendation to finance the third volume of Pareto if feasible.

It was also VOTED to authorize the payment of the editorial services involved in preparing the foreign survey manuscripts for publication.

Committee on the Selection of the Editor for the A.E.R. (Milton Friedman). The Chairman reported that the Committee was considering possible candidates for the position and hoped to announce the name of the new editor by late spring.

Index of Economic Essays and Index of Economic Journals. Richard Ruggles reported that the manuscript for the *Index of Economic Essays,* covering articles published during 1964–65, was ready to be sent to the printer. He recommended that the Association authorize the preparation of an *Index of Economic Essays,* covering articles published in 1966–67 and Volume VIII of the *Index of Economic Journals,* covering the years 1966–67. Following a discussion, it was VOTED to approve an estimated budget of $19,000 to cover the costs of preparing both index volumes. It was VOTED to ask Mrs. Truus Koopmans to continue her editorial work on both volumes.

Committee on Honors and Awards (James Tobin). The Chairman reported by letter the name of the candidate recommended by the Committee for the Distinguished Fellow award.

Nominating Committee (J. J. Spengler). The Executive Committee and the Nominating Committee met as an Electoral College to consider the nominees for the office of president-elect for 1969. After discussion, the nominee was selected and his acceptance obtained. Nominations for the other offices were discussed. The Chairman also submitted the recommendation of the Honors and Awards Committee for the Distinguished Fellow award. The recipient was selected as a result of a ballot of the members of the Electoral College.

10. *Unfinished and Miscellaneous Business.*

History of the Association. Following a discussion, it was VOTED to make a grant of $1,000 to A. W. Coats for the purpose of employing a research assistant to work on the history of the Association.

Committee on Art Exhibits at the Annual Meetings. It was VOTED to authorize the President to appoint a committee to arrange for art exhibits at the annual meetings.

A.E.R. Editorial Office. It was VOTED to make a gift of $1,500 to Miss Doris Merriam on retirement in appreciation of her services as assistant to the Editor of the *American Economic Review.*

Position of the A.E.A. on Questions of Public Policy. President Boulding opened the discussion by noting that at the Business Meeting of the Association held in Washington, D.C., on December 30, 1967, two speakers from the floor had suggested that the Association should, one, consider taking an official stand on the Vietnam war and, two, poll its members regarding their attitudes toward the war. After referring to the provision in the Charter according to which "the Association as such will take no partisan attitude, nor will it commit its members to any position on practical economic questions," the President had agreed to have the Executive Committee consider these suggestions at its next meeting.

In respect to the suggestion that the Association should take an official stand on the Vietnam war, it was noted that this could be done only by changing the Charter. Such a change, in the judgment of the members of the Executive Committee, would be most inadvisable. It was accordingly VOTED to reaffirm strongly the wisdom of the founders of the Association in adopting the provision in the Charter prohibiting the Association as such from taking an official stand or committing its members to any position on economic questions.

In discussing the possibility of polling the Association members in regard to their attitude toward the Vietnam war, the members of the Executive Committee generally agreed that carefully designed and administered opinion polls on policy problems involving the technical competence of economists need not violate the letter or spirit of the Association's Charter. This was, in fact, the position taken at the January 22–23, 1944, Business Meeting of the Association, where a resolution was passed authorizing the taking of polls the objective of which was "to focus the informed opinion and make effective the technical competence of the members of the Association" in matters of economic policy, such as "fiscal policy in war and peace, international monetary proposals, the control of industries affected with public interest, the general relations of government and industry, unemployment, and similar matters." The polls actually taken under the resolution of January 22–23, 1944, were among economists thought to have special competence in respect to (*a*) the proper role of government in the postwar American economy, (*b*) to monetary policy, and (*c*) to agricultural policy.

Reactions regarding the usefulness of the polls taken in 1944 varied widely among members of the profession. At best, in the judgment of various members of the present Executive Committee, the 1944 polls illustrated the practical impossibility of framing relevant questions in such a way that all economists answering them will do so on identical (or even similar) implied assumptions.

Because of these considerations, the Executive Committee VOTED that the Association should not attempt to administer or promote polls on questions of economic policy and that even less should the Association attempt to administer or promote polls on political questions not involving the professional competence of economists.

Program for the 1968 Meeting. President-elect W. J. Fellner discussed his plans for the 1968 meeting.

The meeting was adjourned at 11:00 a.m. on March 9, 1968.

2. Minutes of the meeting held in Chicago, Illinois, December 27, 1968:

The *second meeting* of the 1968 *Executive Committee* was called to order at 2:25 p.m. in the Pick-Congress Hotel, Chicago, Illinois, December 27, 1968. The following were present: K. E. Boulding, presiding, K. F. Arrow, G. H. Borts, J. M. Buchanan, Otto Eckstein, W. J. Fellner, Milton Friedman, J. G. Gurley, W. W. Heller, Fritz Machlup, Mark Perlman, and H. F. Williamson. Absent were: L. R. Klein, Franco Modigliani, Anthony Scott, Arthur Smithies, and R. M. Solow. Present as guests were: J. S. Bain, M. J. Bowman, C. W. Farnham, B. F. Hoselitz, Wassily Leontief, W. F. Owen, Richard Ruggles, and C. L. Schultze.

1. *President's Remarks* (K. E. Boulding). President Boulding outlined the items on the agenda to be covered at the meeting.

2. *Minutes.* The minutes of the March 8–9, 1968, meeting were approved. It was VOTED to hold the spring meeting of the Executive Committee in New York City, March 7–8, 1969, at the New York Hilton Hotel.

3. *Report of the Secretary* (H. F. Williamson). The Secretary announced that George W. McKinney, Jr., of the Irving Trust Company, had agreed to be Chairman of the Committee on Local Arrangements for the 1969 annual meeting to be held in New York City, December 28–30. The headquarters hotel will be the New York Hilton.

Plans for future meetings of the Association were reviewed. At the recommendation of the Secretary, it was VOTED to plan to hold the 1971 meeting in New Orleans and the 1974 meeting in San Francisco. It was also VOTED to reconsider the scheduled meeting in Chicago in 1972 and to explore the possibility of Montreal or Toronto for the 1975 meeting.

The Secretary reviewed other activities of the Association, which are treated in the Secretary's Report given below.

4. *Report of the Treasurer* (H. F. Williamson). The Report of the Treasurer covering the fiscal year 1967 and published below, was summarized. Special attention was called to the fact that the net operating income for the thirteen-month period ending December 31, 1967, was $34,222 and the unappropriated surplus was $433,081. The Treasurer pointed out that it proved to be quite expensive to calculate the undistributed earnings on the securities held by the Association as provided under the accounting system authorized by the Executive Committee on December 26, 1966. Following a discussion, it was VOTED to recommend a change in accounting procedures as outlined in the Report of the Committee on Financial and Investment Policies, printed below.

5. *Report of the Finance Committee* (D. W. Farnham). Mr. Farnham presented the Report of the Finance Committee, published in full below. It was VOTED to accept the report and to extend a special vote of thanks to Mr. Farnham for his services as Chairman of the Finance Committee. It was also VOTED to enter into a formal arrangement with Stein, Roe & Farnham to handle the Association's securities account.

6. *Report of the Committee on Financial and Investment Policies* (Milton Friedman). The report made by the Chairman is printed in full below.

7. *Report of the Managing Editor of the American Economic Review* (J. G. Gurley). The Editor reviewed the operations of the Editorial Office of the *American Economic Review*, covered in full below. It was VOTED to accept the report of the Editor, including the nominations for the Editorial Board, with the understanding that the members will be appointed for staggered terms. On the recommendation of the Editor, it was VOTED to give a bonus of $300 to Joyce Parker for her services to the Editorial Office. It was further VOTED to extend a special vote of thanks to J. G. Gurley for his exceptionally valuable services as Editor of the *American Economic Review.*

8. *Report of the Editor-elect of the American Economic Review* (G. H. Borts). The Editor-elect reviewed future plans for the *American Economic Review.* He expressed his special appreciation to J. G. Gurley and the editorial staff at Stanford for their valuable assistance in transferring the editorial operations to Brown University. He also noted a decision to change the page size of the *Review* from 6 × 9 to 7 × 10 inches and to replace the single-column format with a double-column format. A major consideration prompting this change was the substantial saving that could be achieved in the costs of paper and printing. After discussion, it was VOTED to authorize the following budget proposed by the Editor-elect to cover the operation of the *Review* office for the year 1969:

Editor's salary	$ 9,000

Assistant to the Editor (including Social Security)	8,416
Secretarial and clerical assistant and proofreading (including Social Security)	6,250
Postage and supplies	2,400
Office equipment	485
Expense of establishing new office	500
	$27,051

9. *Report of the Editor of the Journal of Economic Abstracts* (Arthur Smithies). The final report of the *Journal of Economic Abstracts* appears below.

10. *Report of the Editor of the Journal of Economic Literature* (Mark Perlman). The Editor reported satisfactory progress in implementing the plans outlined at the March 8–9, 1968, Executive Committee meeting for the publication of the *Journal of Economic Literature.*

11. *Report of Standing and Special Committees.*

Committee on Research and Publications (I. B. Kravis). In the absence of the Chairman, B. F. Hoselitz reported that, with the exception of the plan to publish the writings of Vilfredo Pareto, work on the "Translations Series" was progressing satisfactorily. The problem in respect to the proposed Pareto volumes came about as a result of a misunderstanding over the nature of the agreement between the Association and Librairie Droz, holder of the copyright to various publications of Pareto, including the *Manuel.* The situation was further complicated by the fact that Augustus M. Kelley has a contract with Librairie Droz to publish the entire translation of the *Manuel.* It was agreed that the Committee on Research and Publications should try to work out a solution of the problem.

Advisory Committee on the Index of Economic Essays (Richard Ruggles). The Chairman reported that Mrs. Truus Koopmans was making excellent progress in the preparation of the manuscript for the *Index of Economic Essays.* At the suggestion of the Chairman, it was VOTED to raise Mrs. Koopmans' annual salary as editor of the series by $300.

Committee on Economic Education (G. L. Bach). The report of the Chairman is printed below. Also printed below is the report of the A.E.A. Visiting Scientists Program, prepared by Phillip Saunders.

N.S.F. Report on the Profession (Arnold Tolles). In the absence of Arnold Tolles, the Secretary noted that the *Studies of the Structure of Economists' Salaries and Income,* prepared by Arnold Tolles and Emanuel Melichar, was issued as a supplement to the December, 1968, issue of the *American Economic Review.*

Advisory Committee on the History of the Association (G. J. Stigler). In the absence of the Chairman, the Secretary reported that A. W. Coats was making good progress on the history of the Association although no date has been set for completion.

12. *Reports of A.E.A. Representatives and Advisory Committees.*

A.C.L.S. (R. L. Andreano). No report.

S.S.R.C. (Franco Modigliani). No report.

N.B.E.R. (W. L. Thorp). Report printed below.

A.A.A.S. (Carl Kaysen). No report.

I.E.A. (Fritz Machlup). Report printed below.

Economics Institute Advisory and Policy Board (Henry Rosovsky). In the absence of Henry Rosovsky, his report, printed below, was outlined by W. F. Owen, who also called attention to the proposal under which the investment funds advanced by the Ford Foundation to the Economics Institute would be managed by the A.E.A. Finance Committee. Following a discussion, it was VOTED to authorize the Finance Committee to work out an arrangement to handle the investment funds of the Institute.

Census Advisory Committee (Solomon Fabricant). Report is printed below.

National Research Council, Division of Behavioral Sciences (J. N. Morgan). Report published below.

13. *Unfinished and Miscellaneous Business.*

Following a discussion of the problems involved in scheduling the 1968 annual meeting in Chicago, the President appointed the following members to an *ad hoc* committee to review the functions of the Association's annual meetings: K. J. Arrow, Chairman, Otto Eckstein, and Franco Modigliani.

The Secretary announced that, in consultation with G. L. Bach and K. E. Boulding, a panel of qualified individuals had been suggested to discuss economics curricula with representatives of a group of Negro colleges at a conference to be held at Fisk University during April, 1969. Funded by the Ford Foundation, the conference is being organized by Dr. Flournoy Coles, Vice-President and Chairman of the Department of Economics at Fisk University.

H. F. Williamson indicated his plan to retire as Secretary-Treasurer on or before the termination of his present term of office, December 31, 1970. He suggested that a committee be appointed to consider the selection of his successor as well as the question of whether it would be necessary or advisable to move the Association's business office from its present location.

The meeting was adjourned at 11:00 p.m.

REPORT OF THE SECRETARY FOR THE YEAR ENDING DECEMBER 31, 1968

Annual Meetings. The final report for the 1967 annual meeting of the Allied Social Science Associations in Washington, D.C., indicated a net registration of approximately 10,800 individuals, of which 45 percent (4,800) were members of the American Economic Association. The net income from the meeting was $38,112. The Association's share, based on the number of members who registered, was $16,992. As an experiment, A.E.A. members were asked to preregister for the meeting. The response was excellent, suggesting that this procedure should be followed at future meetings.

The schedule for future annual meetings is: 1969, New York, New York Hilton; 1970, Detroit, Sheraton-Cadillac; 1971, New Orleans, Jung; 1972, Chicago, Conrad Hilton (tentative); 1973, New York, New York Hilton; 1974, San Francisco, San Francisco Hilton; 1975, Chicago, Conrad Hilton; 1976, Philadelphia (tentative).

Dr. George W. McKinney, Jr., of the Irving Trust Company, New York, has agreed to serve as Chairman of the Committee on Local Arrangements for the 1969 annual meeting.

Membership. Exhibit I below shows that as of December 31, 1968, the total number of members and subscribers was 24,713, a net increase of 1,408 for the year.

Handbook. Plans are well advanced for the publication of a new edition of the "who's who" type of *Handbook* late in 1969.

Advertising and Announcements. There were 184 1/2 pages of paid advertising and 54 pages of exchange advertising in the *American Economic Review* and the *Papers and Proceedings* compared with 192 paid and 50 1/2 exchange pages in 1967. Advertising income for 1968 was $48,794 compared to $52,811 in 1967.

Papers and Proceedings. A comparison of the size and cost of the *Papers and Proceedings* for the period 1963–68 is shown in Exhibit II.

Permission to Reprint and Translate. Official permission to quote from, reprint, or translate and reprint articles from the *American Economic Review* and the *Papers and Proceedings* totaled 155 in 1968 compared to 234 in 1967. Some 42 permissions granted in 1968 were in response to requests to reproduce or reprint articles for classroom or conference use.

Use of the Mailing List. The continued policy of the Association is to grant requests for the use of the mailing list only if the material to be distributed appears to be of interest to a substantial proportion of our members. Net income from the sale of the mailing list during 1968 was $12,008 compared to $9,418 for 1967.

National Register of Scientific and Technical Personnel. Under the contract with the National Science Foundation, the Association in 1968 participated for the third time in the National Register of Scientific and Technical Personnel. Questionnaires were mailed to 22,349 individuals drawn from a variety of lists of economists. The total number of questionnaires returned was 13,940. Of this number, 11,696 met the qualifications for inclusion in the Register as economists and their questionnaires were sent to the National Register Records Center for further processing.

Foreign Students Screening Project. Under the Foreign Student Screening Project, the Association continued during 1968 to answer requests for evaluations of candidates from the major countries in the Middle East, Far East, Latin America, and Africa who applied for admission as graduate students to American and Canadian departments of economics and agricultural economics. The present plan is to discontinue this service at the end of the 1968–69 academic year, with the possibility that the service may be furnished in part by the Economics Institute located at the University of Colorado. Plans are well under way to publish a revised edition of the *Guide to Graduate Study in Economics* in the United States.

Employment Service. According to the final report on the operation of the U.S. Employment Service at the 1967 annual meeting in Washington, D.C., the number of job openings filed was 2,083 and the number of registered applicants was 1,987.

Group Flights. Because the 1968 meeting was held over the weekend and members were automatically entitled to an excursion rate, no attempt was made to schedule group flights to Chicago. An arrangement was made, however, whereby members could be guaranteed space on flights to and from Chicago by purchasing tickets through the Drake Travel Service.

Standing Committees

Ad Hoc ADVISORY COMMITTEE FOR THE JOURNAL OF ECONOMIC LITERATURE

Milton Friedman, *Chairman*
John G. Gurley
Richard H. Leftwich
Richard Ruggles
Arthur Smithies
Harold F. Williamson

ADVISORY COMMITTEE ON THE STUDY OF THE STRUCTURE OF THE PROFESSION

Karl A. Fox, *Chairman*
William J. Baumol
Gary S. Becker
Francis M. Boddy

ADVISORY COMMITTEE ON THE HISTORY OF THE ASSOCIATION

George J. Stigler, *Chairman*
Joseph Dorfman
Harold F. Williamson

CENSUS ADVISORY COMMITTEE

Solomon Fabricant, *Chairman* (1968)
Daniel Creamer (1968)
H. Gregg Lewis (1968)
John Lintner (1968)
Arthur M. Okun (1968)
Morris A. Adelman (1969)
Fred H. Klopstock (1969)
Robert R. Nathan (1969)
Douglass C. North (1969)
Ernest W. Williams (1969)
Gary Fromm (1970)
Guy H. Orcutt (1970)
Donald J. Daly (1971)
Bert Hickman (1971)
Sherman J. Maisel (1971)
Murray L. Weidenbaum (1971)

COMMITTEE ON THE AMERICAN ECONOMIC REVIEW AND THE JOURNAL OF ECONOMIC ABSTRACTS

Milton Friedman, *Chairman*
Kenneth E. Boulding
Bernard F. Haley
Neil H. Jacoby
Fritz Machlup
Robert H. Strotz
Harold F. Williamson, Ex Officio

COMMITTEE ON ECONOMIC EDUCATION

G. L. Bach, *Chairman* (1970)
Marshall Colberg (1968)
Rendigs Fels (1968)
Kenneth E. Boulding (1969)
Allen C. Kelley (1969)
Henry H. Villard (1970)
Harold F. Williamson, Ex Officio

COMMITTEE ON HONORS AND AWARDS

James Tobin, *Chairman* (1970)
Richard A. Musgrave (1970)
Anthony D. Scott (1970)
William H. Nicholls (1972)
Douglass C. North (1972)
George P. Shultz (1972)

COMMITTEE ON RESEARCH AND PUBLICATIONS

Irving B. Kravis, *Chairman* (1969)
Bert F. Hoselitz (1968)
Franco Modigliani (1968)
Albert Rees (1969)
Zvi Griliches (1970)
Willard L. Thorp (1970)

COMMITTEE ON ADMINISTRATIVE STAFF SALARIES

Milton Friedman
Kenneth E. Boulding
William J. Fellner

ECONOMICS INSTITUTE POLICY AND ADVISORY BOARD

Henry Rosovsky, *Chairman* (1971)
Neil W. Chamberlain (1969)
H. Gregg Lewis (1969)
Douglass C. North (1970)
Vernon W. Ruttan (1970)
Joseph J. Spengler (1971)
Wyn F. Owen, Ex Officio

JOINT COMMITTEE WITH THE ASSOCIATION OF AMERICAN LAW SCHOOLS

Armen A. Alchian
Ronald H. Coase
Mark S. Massel
Gerald M. Meier
George J. Stigler

JOINT COUNCIL ON ECONOMIC EDUCATION, SPECIAL ADVISORY COMMITTEE

G. L. Bach (1970)
Rendigs Fels (1968)
Ben W. Lewis (1969)

TECHNICAL SUBCOMMITTEE ON BUSINESS CYCLE DEVELOPMENTS

Geoffrey H. Moore (1968)
Arthur K. Okun (1968)
Frank E. Morris (1969)
Harry Stark (1969)
Gottfried Haberler (1970)
Bert G. Hickman (1970)
Donald J. Daly (1971)
Gary Fromm (1971)
Beryl W. Sprinkel (1971)
Kenneth B. Williams (1971)

Committees Appointed During the Year

COMMITTEE ON ELECTIONS

Robert Eisner, *Chairman*
Walter D. Fisher
Harold F. Williamson

FINANCE COMMITTEE

C. Wells Farnham, *Chairman*
Corliss D. Anderson
Milton Friedman
Harold F. Williamson

NOMINATING COMMITTEE

Joseph J. Spengler, *Chairman*
John M. Letiche
Ruth P. Mack
Robert V. Roosa
Phillip Saunders
Robert A. Solo

Council and Other Representatives

A.A.A.S.

Carl Kaysen (1968)

A.C.L.S.

Ralph L. Andreano (1970)

ARCHIVES ADVISORY COUNCIL, GENERAL SERVICES ADMINISTRATION

Harold F. Williamson

CONSORTIUM OF PROFESSIONAL ASSOCIATIONS FOR THE STUDY OF TEACHER IMPROVEMENT PROGRAMS

G. Heberton Evans
Harold F. Williamson

I.E.A.

John T. Dunlop (1968)
Fritz Machlup (1970)

N.B.E.R.

Willard L. Thorp (1970)

NATIONAL COUNCIL OF TESTING OF ENGLISH AS A FOREIGN LANGUAGE

Arthur E. Burns

NATIONAL RESEARCH COUNCIL, DIVISION OF BEHAVIORAL SCIENCES

Harold J. Barnett (1969)
Lionel W. McKenzie (1970)
Guy H. Orcutt (1971)

ASSOCIATED ORGANIZATIONS FOR TEACHER EDUCATION

Henry H. Villard

COMMITTEE ON QUANTITATIVE DATA

Richard A. Easterlin

S.S.R.C.

Franco Modigliani (1968)
Albert Rees (1969)
Zvi Griliches (1970)

Representatives of the Association on Various Occasions

OREGON STATE UNIVERSITY CENTENNIAL
Marshall D. Wattles

COLLEGE OF THE HOLY CROSS 125TH ANNIVERSARY CONVOCATION
Robert Cheney

OKLAHOMA STATE UNIVERSITY, NATIONAL MANPOWER CONFERENCE
Richard H. Leftwich

ROCHESTER INSTITUTE OF TECHNOLOGY, DEDICATION OF NEW CAMPUS
Lionel W. McKenzie

BENTLEY COLLEGE, DEDICATION OF NEW WALTHAM CAMPUS
Paul R. Nichols

INAUGURATIONS
Robert Edwin Kennedy, California State Polytechnic College
Alec P. Alexander
Richard H. Timmins, Huron College
Robert Beck
Luther Wesley White, III, Randolph-Macon College
Thomas S. Berry
Morris Berthold Abram, Brandeis University
Robert L. Bishop
Robert Morse, Case Western Reserve University
Diran Bodenhorn
Robert Owens, Knoxville College
Roger L. Bowlby
Paul Russell Anderson, Temple University
Dorothy S. Brady
Irving Carlson, Westminster College
Jacob Cohen
W. Seavey Joyce, Boston College
Arthur H. Cole
Brage Golding, Wright State University
Paul G. Craig
Allen Keith Jackson, Huntingdon College
Dale Cramer
Frederick C. Davison, University of Georgia
James Crawford
George Howard Williams, American University
Clarence H. Danhof
John Arthur Peoples, Jr., Jackson State College
J. Anderson Davis
Ferrel Heady, University of New Mexico
Henry C. Dennis
Marvin August Rapp, Onondaga Community College
Melvin A. Eggers
Woodrow W. Stickler, University of Louisville
Carl C. Erwin
James Raymond Lawson, Fisk University
Rendigs T. Fels
Samuel E. Braden, Illinois State University
Marvin Frankel
Martha Peterson, Barnard College
C. Lowell Harriss
Walter Scott Johnson, Spokane Community College
Millard Hastay
Isaac Henry Miller, Jr., Bennett College
Milton S. Heath
Kurt Richard Schmeller, Queensborough Community College of the City U. of N.Y.
Robert A. Kavesh
James Grindling Harlow, West Virginia University
Forest H. Kirkpatrick
Roland Dille, Moorhead State College
W. E. Koenker
Emerson C. Shuck, Eastern Washington State College
Frank J. Kottke
Robert Fleming, University of Michigan
Robert F. Lanzillotti
Edward Hirsch Levi, University of Chicago
Simeon Leland
James Britt Donovan, Pratt Institute
Edward Marcus
Theodore David Lockwood, Trinity College
Richard A. Miller
Samuel E. Stumpf, Cornell College
Paul R. Olsen
Albert N. Whiting, North Carolina College
Ralph W. Pfouts
James O. Fuller, Fairleigh Dickinson University
Sidney Ratner
John J. Pruis, Ball State University
Charles Leese
E. K. Fretwell, Jr., State University of N.Y. at Buffalo
Richard N. Rosett
William Carter, Wisconsin State University
Richard Ruth
Milligan Maceo Nance, Jr., South Carolina

State College
Alfred G. Smith, Jr.
Benjamin F. Payton, Benedict College
Alfred G. Smith, Jr.
Charles B. Kinney, Mattatuck Community College
William P. Snavely
Werner A. Baum, University of Rhode Island
Jerome L. Stein
Lloyd Vincent, Angelo State College
Carey C. Thompson
Norman A. Wiggins, Campbell College
W. D. Toussaint
Charles Johnston Hitch, University of California at Los Angeles
Daniel C. Vandermeulen
Hugh Morris Gloster, Morehouse College
William T. Whitman
George Reed Field, Wisconsin State University
Robert E. Will
Malcolm Moos, University of Minnesota
Robert E. Will
Stanley John Idzerda, College of Saint Benedict
William V. Williams

Harold F. Williamson, *Secretary*

EXHIBIT I

Members and Subscribers

	Totals 12/31/67	Gain or *Loss*	Totals 12/31/68
Class of membership:			
Annual	14,017	1,068	15,085
Junior	2,031	40	2,071
Family	226	22	248
Complimentary	122	24	146*
Life	262	6	268
Honorary	17	—	17
Total members	16,675	1,160	17,835
Subscribers	6,630	248	6,878
Totals	23,305	1,408	24,713

* Includes 75 who do not receive the publications.

EXHIBIT II

Publication Costs

	Papers and Proceedings			Handbooks		
Year*	Number of pages	Number of copies	Cost	Number of pages	Number of copies	Cost
1957	754	12,400	$16,253	548	10,100	$15,815
1958	677	12,700	15,471	32	9,300	1,434
1959	689	14,000	16,780			
1960	745	14,800	18,914			
1961	675	15,800	19,759			
1962	615	16,000	18,277			
1963	753	17,700	23,440			
1964	698	18,500	23,362	472	16,000	48,626
1965	652	20,000	23,264			
1966	670	22,500	28,405	184	23,000	13,261
1967	741	24,000	32,682			
1968	751	26,500	36,629			

* This is the year of publication and pertains to the meeting of the preceding year.

REPORT OF THE TREASURER FOR THE THIRTEEN-MONTH PERIOD ENDING DECEMBER 31, 1967

The financial result of the operations of the Association for the thirteen-month period ending December 31, 1967, is shown in detail in the Auditor's Report, Exhibits 3 and 4, printed below. Under the system of accounting adopted as of December 26, 1966, income exceeded expenses for this period by a total of $34,222. Under the previous method of accounting, this amount would have been $26,887.

A detailed account of the Association's financial status for 1967 is also shown in the Auditor's Report, Exhibit 1. As indicated, the general fund balance of unappropriated surplus of the Association at the end of 1967 was $433,081.

Tables 1 and 2 give summaries of the investment portfolios and return on investment for the years 1945–67. A list of the Association's security holdings for 1968 is shown in Table 1 of the Report of the Finance Committee, printed below.

Harold F. Williamson, *Treasurer*

TABLE 1

Investment Portfolio

Year	At Par	Cost			Market
	Bonds	Bonds	Stocks	Total	Stocks and Bonds
1945	$ 40,000	$ 36,705	$ 44,955	$ 81,661	$103,574
1948	35,000	33,108	48,624	81,732	84,841
1950	35,000	33,108	51,978	85,087	104,177
1951	43,000	43,340	49,764	93,104	117,316
1952	42,000	42,312	58,934	101,246	130,836
1953	68,000	68,308	46,458	114,766	134,562
1954	61,000	61,518	38,082	99,600	132,280
1955	75,000	75,370	59,394	134,764	166,772
1956	75,000	75,370	60,237	135,607	168,337
1957	75,000	75,370	55,084	130,454	151,638
1958	75,000	75,370	67,741	143,111	175,609
1959	75,000	75,386	67,652	143,038	191,506
1959*	175,000	175,616	67,652	243,268	291,506
1960*	160,000	160,508	94,910	255,418	299,768
1961*	170,000	169,794	109,071	278,865	356,131
1962*	125,000	125,367	116,699	242,066	293,039
1963*	90,000	90,367	119,524	209,891	284,160
1964*	150,000	149,198	131,249	280,447	371,556
1965*	125,000	124,048	157,836	281,884	409,653
1966*	143,000	142,138	241,075	383,213	491,324
1967	110,000	109,799	249,672	359,471	546,024

* Includes bonds held in temporary operating fund.

TABLE 2

Return on Investments

Year	Bonds	Stocks	Total	Rate of Return on Cost
1945	$1,479	$2,488	$3,968	4.71%
1948	1,194	2,944	4,139	5.06
1950	1,117	3,860	4,977	5.85
1951	1,026	4,607	5,633	6.05
1952	1,117	3,681	4,799	4.75
1953	1,435	3,587	5,022	4.36
1954	1,621	2,961	4,582	4.58
1955	1,750	3,002	4,752	3.53
1956	1,770	3,336	5,106	3.76
1957	1,770	3,397	5,167	3.90
1958	1,770	3,182	4,952	3.46
1959*	2,518	3,231	5,749	3.90
1959†	3,894	3,231	7,125	2.90
1960†	6,693	3,772	10,465	4.09
1961	5,460	4,143	9,603	3.44
1962	4,838	4,489	9,327	3.85
1963‡	3,320	5,041	8,361	3.98
1964‡	3,341	6,364	9,705	3.46
1965‡	5,286	6,330	11,616	4.12
1966‡	6,356	7,957	14,312	3.73
1967	7,906	11,434	19,340	5.38

* Does not include income from bonds held in temporary operating fund.

† Includes income fron bonds held in temporary operating fund.

‡ Does not include interest on savings account.

1. *Present method of reporting income from equities:* Currently (pre-1967) income from equities was reported as the sum of dividends received and capital gains realized.

2. *Defects of present method:* This method has two main defects. First, it does not recognize the market's evaluation of changes in the future prospects of the enterprises whose stocks we own. Second, it makes income depend on the accident of when securities are sold. The first principle might lead to either understatement or overstatement, but in practice has led to understatement. The second produces wide fluctuations in recorded income that do not necessarily correspond to changes in the amount available for expenditures without impinging on capital.

3. *Proposed alternative:* We recommend remedying the defects as follows:

(3.1) *Changes in prospects of enterprises:* To eliminate the first effect, we propose computing for each year the changes in the market value of the stocks in our portfolio adjusted for changes in the general price level. The adjustment will be made by subtracting the product of the start of the year market values of the portfolio times the percentage change in a price index. We propose to use as the price index the implicit GNP deflator, and to measure the percentage change from fourth quarter to fourth quarter. However, the adjusted change in market value computed in this way will not be entered in full each year in order to overcome the second defect discussed in the next paragraph.

(3.2) *Realized capital gains and associated arbitrary fluctuations:* By recognizing as in the previous paragraph, all changes in market value, we would remove the fortuitous element in our present method arising from the timing of sales. But, including all such gains would give undue significance to short-term market fluctuations. Since the objective is to take account of longer-term movements in stock prices, we propose spreading the gain as calculated in the previous paragraph over three years, including one-third in the year to which the calculation refers and one-third in each of the next two years. The number "three" was selected as roughly corresponding to the duration of cycles in stock market prices in the post-World War II period.[1]

(3.3) *Final formula:* This treatment would make the Association's income from equities in each year equal to dividends received in that year plus one-third of market appreciation in that year (adjusted for price-level change) plus one-third of adjusted appreciation in the preceding year, plus one-third of adjusted appreciation in the second preceding year.

MILTON FRIEDMAN, *Chairman*

---

[1] In the post-World War II period, peaks in stock prices were reached in 1946, 1948, 1953, 1956, 1959, 1961 and 1966. There were six cycles in the 20-year period, or an average duration of 3⅓ years. In the post-World War I period, 1919 to 1939, there were only six peaks, or an average duration of 4 years.

REPORT OF THE FINANCE COMMITTEE

Executive Committee
American Economic Association
Evanston, Illinois

Gentlemen:

The accompanying inventory lists the securities held by the American Economic Association as of December 13, 1968, with the costs and the market values as of December 13. The total market value of the securities account as of December 31, 1967, was $544,772. Value as of December 13, 1968, was $638,763. We estimate that the market value increased by 17.27 percent. As a basis of comparison, the Dow-Jones Industrial averages increased by 8.41 percent and the Standard & Poor's average of 425 stocks increased by 11.56 percent.

During the current year, there were no important changes in the bond portion of the investment account. A number of common stocks were sold and purchased with no important change in the aggregate common stock investment. Among issues sold were 200 Marsh & McLennan, 200 Olin Mathieson, 1,000 Jostens and 134 Squibb Beech-Nut. These were sold at a profit of about $32,000 and the proceeds invested in 500 Public Service Company of Northern Indiana, 200 Dow Chemical, and 100 additional Deere & Company, and 180 LaSalle National Bank. Other changes during the year were 2-for-1 stock splits by St. Paul Fire & Marine, Atlantic Richfield, Marathon Oil, and International Business Machines. International Nickel has split 2½-for-1 and Royal Dutch declared a 10 percent stock dividend.

At this time, the undersigned believes it is appropriate to consider the sale of some common stock on balance recognizing that long-term investment policy is to be aggressive. Also, consideration should be given to the purchase of more long-term bonds, whether or not there is a reduction of common stocks.

C. Wells Farnham, *Chairman*

TABLE 1

Inventory and Appraisal of Securities and Cash as of 12/13/68

	Par or Shares	Market Value	Cost
CASH AND SHORT-TERM OBLIGATIONS			
CASH		0	
Governments			
U.S. Treas. 3⅞ 11/15/71	20,000	18,800	19,951
U.S. Treas. 2½ 12/15/72–67	15,000	13,350	14,456
U.S. Treas. 5¾ 2/15/75	25,000	24,500	25,013
TOTAL CASH AND SHORT-TERM SECURITIES		56,650	59,420
Long-Term Bonds and Preferred Stocks			
Anheuser Busch 6 7/1/92	10,000	8,900	10,000
Corn Products 5¾ 8/15/92	10,000	8,800	9,950
General Tel. 6¼ 12/1/91	20,000	17,800	20,530
Pet Inc. 5⅞ 6/1/92	10,000	8,800	9,900
TOTAL LONG-TERM BONDS AND PREFERRED STOCKS		44,300	50,380
TOTAL CASH AND FIXED INCOME SECURITIES		100,950	109,800
Convertible Securities			
Preferred Stocks			
Union Oil 2.50 CV. Pfd.	200	16,800	12,890
General American Transportation 2.50 CV. Pfd.	400	25,200	13,862
TOTAL CONVERTIBLE SECURITIES		42,000	26,752
COMMON STOCKS			
Utilities			
Central and South West	300	13,200	2,100
Houston Ltg. and Pwr.	200	9,000	827
Peoples Gas Co.	275	11,275	3,560
Public Service Co. of No. Indiana	500	23,000	24,215
		56,475	30,702

TABLE 1 *(Continued)*

	Par or Shares	Market Value	Cost
Financial			
First Bank System	400	24,000	12,690
St. Paul Fire and Marine	400	17,600	12,640
Wells Fargo Bank	300	18,900	10,241
		60,500	35,571
Paper and Forest Products			
Inland Container	150	6,150	4,944
Machinery and Construction			
Deere and Co.	300	16,500	9,484
Rex Chain Belt	300	13,100	6,620
		29,700	16,104
Transportation			
Eastern Air Lines	200	6,200	8,950
United Air Lines	300	12,300	14,741
		18,500	23,691
Mining and Metals			
International Nickel	250	9,250	3,911
Mc Intyre Porcupine Mines	100	8,500	4,818
		17,750	8,729
Oil and Gas			
Atlantic Richfield	200	24,800	9,625
Marathon Oil	200	11,000	5,630
Royal Dutch	440	23,760	15,955
Standard Oil of Indiana	200	12,600	3,650
		72,160	34,860
Chemicals and Drugs			
Abbott Labs	300	22,400	6,135
Dow Chemical	200	15,800	16,119
		38,200	22,254
Electrical Products			
Motorola	150	19,950	6,745
Westinghouse Electric	200	14,400	11,615
Zenith Radio	300	17,400	1,975
		51,750	20,335
Office Equipment			
IBM	122	38,674	10,169
Xerox	50	13,650	5,305
		52,324	15,474
Miscellaneous			
Disney Productions	454	41,314	8,356
R. R. Donnelley	300	14,100	9,795
Marcor	150	8,250	5,013
		63,664	23,164
Foreign			
Bayer A. G.	150	7,500	6,820
Siemens A. G.	110	8,140	5,173
Swedish Ball Bearing	200	13,000	14,525
		28,640	26,518
TOTAL COMMON STOCKS		495,813	262,346
TOTAL SECURITIES AND CASH		638,763	398,898

STOCK DIVIDENDS AND OTHER CHANGES SINCE 10/31/68
Withdrawal 50M Treasury Bills due 12/26/68 to Operating Funds.
Change of name from Montgomery Ward to Marcor.
Exchange of 400 La Salle National Bank to 400 General American Transportation $2.50 CV. Pfd.

REPORT OF THE AUDITOR

To the Executive Committee of
American Economic Association:

We have examined the statement of assets and liabilities of American Economic Association (a District of Columbia corporation, organized not for profit) as of December 31, 1967, and the related statements of changes in restricted fund balances and income and expenses for the thirteen-month period then ended. Our examination was made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

In our opinion, the accompanying statements present fairly the assets and liabilities of American Economic Association as of December 31, 1967, and the changes in restricted fund balances and income and expenses for the thirteen-month period then ended, in conformity with generally accepted accounting principles applied, except for the change (which we approve) in accounting for investments as described in the note to the financial statements, on a basis consistent with that of the preceding year.

ARTHUR ANDERSEN & Co.

Chicago, Illinois,
February 22, 1968.

EXHIBIT 1

AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF ASSETS AND LIABILITIES
DECEMBER 31, 1967

ASSETS		
CASH		$ 94,649
INVESTMENTS (see note):		
U.S. Government obligations and corporate bonds, at market value (cost of $109,799)	$104,575	
Corporate stocks, at market value (cost of $249,672)	441,449	546,024
ACCOUNTS RECEIVABLE, less reserve of $1,600		34,484
ADVANCES, relating to United States government grant for foreign student screening and selection		6,698
PREPAID EXPENSES		756
OFFICE FURNITURE AND EQUIPMENT, at cost, less reserve for depreciation of $2,793		4,385
		$686,996

LIABILITIES		
ACCOUNTS PAYABLE		$ 35,542
RESERVE FOR HANDBOOK		32,000
DEFERRED INCOME:		
Life membership dues	$ 30,000	
Prepaid membership dues	15,103	
Prepaid subscriptions	36,704	81,807
DEFERRED INCREMENT IN MARKET VALUE OF INVESTMENTS (see note)		46,489
RESTRICTED FUND BALANCES (Exhibit 2)		58,077
GENERAL FUND:		
Balance, November 30, 1966, as previously reported	$267,628	
Add—Adjustment to state investments at market value at November 30, 1966 (see note)	131,231	
Balance, November 30, 1966, as adjusted	$398,859	
Add—Excess of income over expenses for the thirteen-month period ended December 31, 1967 (Exhibit 3)	34,222	433,081
		$686,996

The accompanying note to financial statements is an integral part of this statement.

NOTE TO FINANCIAL STATEMENTS
DECEMBER 31, 1967

Effective December 1, 1966, the Association changed from a cost to a market value basis of accounting for investments. Under the new method, that portion of the increment in market value of corporate stocks which is equivalent to the increase in the underlying retained earnings of those stocks is reflected in income currently and the balance of the increment, after adjusting for an inflation factor, is recognized in income over a three-year period. The change in market value of U. S. government obligations and corporate bonds, after adjusting for an inflation factor, is reflected in income currently.

This change in accounting for investments had the effect of increasing the excess of income over expenses for the thirteen months ended December 31, 1967, by $8,833 which is arrived at as follows:

Increment in market value of corporate stocks recognized in current year income		$26,887
Deduct—		
Decline in market value of U.S. government obligations and corporate bonds	$ 6,900	
Gain on sale of securities	11,154	18,054
Net increase		$ 8,833

EXHIBIT 2
AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF CHANGES IN RESTRICTED FUND BALANCES FOR THE THIRTEEN-MONTH PERIOD ENDED DECEMBER 31, 1967

Description of Restricted Funds	Balance Nov. 30, 1966	Receipts	Disbursements	Balance Dec. 31, 1967
The Ford Foundation grants for—				
Translation of foreign economic publications and survey of foreign economic research	$35,165	$ —	$ (10,120)	$25,045
Clearinghouse for economists interested in overseas assignments	2,049	52	(2,101)	—
Foreign student screening and selection	35,107	959	(15,879)	20,187
United States government grant for foreign student screening and selection	9,935	—	(2,666)	7,269
The Asia Foundation for Asian economist membership dues to the American Economic Association and related travel expenses	2,514	2,500	(2,824)	2,190
The Agricultural Development Council, Inc.	320	—	(320)	—
The National Science Foundation grant for visiting scientist program	1,827	6,717	(5,387)	3,157
Sundry grants	155	200	(126)	229
	$87,072	$10,428	$ (39,423)	$58,077

EXHIBIT 3

AMERICAN ECONOMIC ASSOCIATION

STATEMENT OF INCOME AND EXPENSES FOR THE THIRTEEN-MONTH PERIOD ENDED DECEMBER 31, 1967

INCOME FROM DUES AND ACTIVITIES:			
Membership dues			$149,553
Subscriptions			66,163
Advertising			52,811
Sale of copies, republications and handbooks			7,904
Sale of mailing list			9,418
Annual meeting			16,564
Sundry			5,871
			$308,284
INVESTMENT INCOME (see note):			
U.S. government obligations and corporate bonds—			
Interest	$ 7,906		
Less—Decline in market value (see note)	6,900	$ 1,006	
Corporate stocks—			
Cash dividends	$11,434		
Increment in market value recognized (see note)	26,887	38,321	39,327
Total income			$347,611
EXPENSES:			
Publication (Exhibit 4)		$214,981	
Administrative (Exhibit 4)		88,408	
Grant to Yale University for *Index of Economic Essays*		10,000	313,389
Excess of income over expenses for the thirteen-month period ended December 31, 1967			$ 34,222

The accompanying note to financial statements and Exhibit 4 are an integral part of this statement.

EXHIBIT 4

AMERICAN ECONOMIC ASSOCIATION

STATEMENT OF PUBLICATION AND ADMINISTRATIVE EXPENSES FOR THE THIRTEEN-MONTH PERIOD ENDED DECEMBER 31, 1967

PUBLICATION EXPENSES:	
Printing and mailing of—	
American Economic Review	$ 81,984
Journal of Economic Abstracts	51,220
Papers and Proceedings	32,682
Provision for *Handbook*	8,000
Addressing machine rental and expenses	3,228
Editorial salaries	31,702
Payments to contributors	2,735
Editorial supplies and expense	3,430
Total publication expenses	$214,981
ADMINISTRATIVE EXPENSES:	
Secretary's salary, less $4,350 allocated to restricted funds	$ 9,750
Office salaries	43,228
Office machine rentals	689
Postage	5,849
Stationery, printing and supplies	3,008
Insurance	505
Annual meeting expenses	4,724
Executive Committee expenses	2,795
Committee on Economic Education expenses	1,369
Other committees' expenses	2,156
International Economic Association dues and expenses	728
Archives	138
Pension annuity payments	1,999
Payroll taxes	2,157
Depreciation of furniture and equipment	571
Telephone and telegraph	1,347
Committee on Research and Publications	1,550
President and president-elect expenses	2,000
Sundry	5,094
Expenses allocated to restricted funds	(1,249)
Total administrative expenses	$ 88,408

AMERICAN ECONOMIC REVIEW

REPORT OF THE MANAGING EDITOR FOR THE YEAR ENDING DECEMBER 1968

The number of manuscripts submitted during 1968 was 637, 103 more than last year and almost 400 more than a decade ago. Table 1 shows these figures for each year since 1948; it also records the number of manuscripts published each year, and the ratios of published papers to the totals received. During the past year, 93 papers were published and only 65 were accepted, reducing once again the large backlog of papers awaiting publication.

TABLE 1

Manuscripts Submitted and Published, 1948–68

Year	Submitted	Published	Ratio of Published to Submitted
1948	205	48	.23
1949	200	53	.27
1950	197	53	.27
1951	222	49	.22
1952	190	47	.25
1953	234	51	.22
1954	231	47	.20
1955	245	41	.17
1956	242	48	.20
1957	215	40	.19
1958	242	46	.19
1959	279	48	.17
1960	276	46	.17
1961	305	47	.15
1962	273	46	.17
1963	329	46	.14
1964	431	67	.16
1965	420	59	.14
1966	451	62	.14
1967	534	94	.18
1968	637	93	.15

Table 2 provides the breakdown of the 1968 volume's contents between articles, review articles, communications, book reviews, etc. Similar information for 1966 and 1967 is also shown. The number of pages devoted to leading articles, review articles, and communications was about the same in total this year as last, but book reviews took up substantially more space.

Table 3 shows the several most popular "fields" in 1968, as reflected in the subject matter of all manuscripts submitted during

TABLE 3

Number of Manuscripts Submitted in the Most Popular Fields in 1968

Microtheory	115
International trade and finance	68
Monetary and banking theory and policy	64
Economics of underdeveloped countries	35
Fiscal and tax theory and policy	29
Labor economics	25
Industrial organization	25

TABLE 2

Summary of Contents, 1966–68

	1967		1966		1968	
	Number	Pages	Number	Pages	Number	Pages
Leading articles	30	539	25	595	29	588
Review articles	3	41	0	0	1	12
Communications:						
Original	24	138	13	83	19	113
Comments and replies	37	178	24	114	44	174
Books reviewed	196	376	210	417	218	424
Classified lists:						
New books		77		71	—	82
Periodicals		—		15	—	—
Dissertations		52		39	—	63
Notes		64		57	—	67
		1465*		1392*		1523

* Includes some blank pages.

TABLE 4

SUBJECT-MATTER DISTRIBUTION OF ACCEPTED MANUSCRIPTS IN 1968

	Articles	Review Articles	Original Communications	Comments; Replies	Totals
General economics	1				1
Price theory	6		4	1	11
Income theory	1		1	2	4
History economic thought					
Economic history					
Economic development; national economies	2		1	2	5
Economic statistics; accounting	2		1	1	4
Money and banking	5		1		6
Public finance	3			2	5
International economics	5		3	2	10
Business finance	1				1
Business organization and administration					
Industrial organization	2		2		4
Land economics; housing, etc.	4				4
Labor economics	3			4	7
Population; welfare, etc.			1		1
Unclassified	1		1		2
	36	0	15	14	65

the year. Almost 40 percent of all manuscripts submitted fell in the three areas of microtheory, international economics, and monetary economics; about 60 percent fell in the seven areas listed above. Other popular fields, not shown in the table, were education, health, and poverty.

Table 4 summarizes the subject matter distribution of accepted articles, review articles, and communications for 1968. The most interesting figures are those for articles and original communications, since to some extent these figures indicate the areas in which most work of publishable quality and of broad interest to economists is being done. The table shows, however, that we have accepted manuscripts in almost all of the listed areas.

Table 5 presents the actual annual expenditures from 1964 through 1968 for the four regular issues of the *Review* and the estimated budgets for 1968 and 1969. Actual expenditures have risen since 1964 owing principally to increases in printing and mailing expenses, which in turn can be traced to the increased number of pages per copy, the larger number of copies printed, and increases in basic printing costs.

Budgeted expenditures for 1968 were $120,626, while actual expenditures were $127,294. The difference is owing mainly to higher printing and mailing costs; my budgeted expenditure for this item was based on an estimate of 106,000 copies of the *Review* for the year and 444 pages per copy, but the actual number of copies was almost 107,000 and the actual number of pages per copy was 457.

Table 6 gives detailed information about printing and mailing costs by issues during 1968. The number of copies printed in 1968 averaged 26,744 per issue compared to 24,750 the year before and to 22,681 in 1966. Ten years ago, the figure was only 12,400. The successive increases are of course mainly related to the increases in membership and subscriptions.

During the year I have had the advice and assistance of the following foreign correspondents, who have been particularly helpful with regard to the selection of foreign books for listing and review:

Maurice Flamant (France)
Erich Schneider (Germany)

Four members of the Board of Editors complete their three-year terms of office at this time: Hollis Chenery, Robert Eisner, Stanley Lebergott, and Richard Nelson. In addition, Warren Smith and Hirofumi Uzawa, for per-

TABLE 5

ACTUAL AND BUDGETED EXPENDITURES, 1964–69

	Actual Expenditures					Budget	
	1964	1965	1966	1967	1968	1968	1969
Printing and mailing...	$53,660.84	$61,606.82	$74,277.93	$82,840.41	$95,414.42	$90,000.00	
Payments to contributors........	1,424.50	1,598.00	1,873.50	2,255.00	2,140.50	2,000.00	
Subtotal..........	$55,085.34	$63,204.82	$76,151.43	$85,095.41	$97,554.92	$92,000.00	
Office Expenses							
Editor's salary........	$ 8,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00
Editorial assistance....	13,460.06	12,024.08	14,189.83	16,218.65	17,850.91	17,246.00	14,666.00
Assistant to editor...	6,619.32	6,619.32	7,330.00	8,025.00	8,465.64	8,426.00	8,416.00
Book reviews.......	1,833.32	2,000.00	2,000.00	2,000.00	2,205.00	2,100.00	—
Secretarial..........	1,632.75	1,725.00	2,160.15	2,400.00	2,644.20	2,520.00	4,000.00
Clerical, proofreading, etc...........	3,374.67	1,679.76	2,699.68	3,793.65	4,536.07	4,200.00	2,250.00
Postage and supplies...	1,348.23	1,462.32	1,479.61	1,895.78	2,888.27	2,280.00	2,400.00
Office equipment......	122.07	—	175.50	—	—	100.00	485.00
Subtotal..........	$22,930.36	$22,486.40	$24,844.94	$27,114.43	$29,739.18	$28,626.00	$26,551.00
Total	$78,015.70	$85,691.22	$100,996.37	$112,209.84	$127,294.10	$120,626.00	—

TABLE 6

COPIES PRINTED, SIZE AND COST OF PRINTING AND MAILING IN 1968

	Printed	Pages		Cost		
		Net	Gross	Issue	Reprints	Total
March......	26,500	342	432	$22,374.91	$210.82	$22,585.73
June........	26,476	338	392	20,611.23	150.90	20,762.13
September ..	27,000	478	540	26,475.61	149.77	26,625.38
December ...	27,000	365	464	22,825.00*	150.00	22,975.00
	106,976	1,523	1,828	$92,286.75	$661.49	$92,948.24

* Estimated.

sonal reasons, resigned earlier in the year. The Association owes all of them a heavy debt of gratitude for the generous expenditure of time they have made in the interests of the *Review*, and I very much appreciate their constant willingness, even when other obligations pressed, to review manuscripts and offer editorial advice. I nominate as their successors on the Board and as additions to the Board, for three-year terms beginning in 1969: Carl Christ, Charles Ferguson, Alvin Marty, Harry Johnson, G. Warren Nutter, Vernon Smith, Herbert Stein, and Lloyd Ulman.

During the year I have frequently sought the aid of members of the profession in addition to the members of the Editorial Board—partly to relieve the latter of what would otherwise be an impossibly heavy burden and partly to obtain advice of specialists in particular areas not represented on the Board. The following have assisted in this way:

M. Abramovitz
I. Adelman
M. Adelman
S. Alexander
A. Ando
K. Arrow
G. Bach
J. Bain
R. Baldwin
B. Balassa
W. Baumol
D. Bear
A. Bergson
J. Berliner
G. Bierwag
R. Bishop

A. Bloomfield
G. Borts
R. Brandis
G. Break
M. Bronfenbrenner
E. Brown
F. Brown
G. Brown
H. Bruton
J. Buchanan
E. Budd
P. Cagan
R. Caves
J. Chipman
G. Chow
C. Christ
R. Clower
R. Coase
R. Coen
J. Coleman
J. Conlisk
R. Cooper
J. Coppack
P. Cootner
R. Crouch
P. David
O. Davis
R. Davis
L. DeAlessi
F. deLeeuw
H. Demsetz
E. Despres
D. Dewey
P. Diamond
J. Dreze
J. Dunlop
R. Easterlin
D. Fand
R. Farmer
E. Feige
C. Ferguson
M. Friedman
G. Fromm
V. Fuchs
L. Gallaway
F. Gehrels
H. Georgiadis
A. Goldberger
S. Goldfeld
S. Goldman
R. Goode
M. Gordon
R. Gordon
M. Gort
P. Gray
R. Gray
M. Greenhut
Z. Griliches
H. Grubel
P. Gutmann
J. Guttentag
R. Hall
D. Hamberg
W. Hansen
A. Harberger
D. Hester
B. Hickman
G. Hildebrand
J. Hirshleifer
F. Holzman
G. Horwich
H. Houthakker
C. Howe
L. Hurwicz
J. Ingram
M. Intriligator
A. Johnson
M. Johnson
A. Kahn
N. Kaplan
J. Kareken
D. Keesing
M. Kemp
P. Kenen
R. Kessel
L. Klein
S. Klein
S. Kolm
M. Kreinin
J. Krutilla
M. Kurz
S. Kuznets
K. Lancaster
L. Lau
G. Lauman
T. Lee
L. Lefeber
R. Leftwich
H. Leibenstein
A. Lerner
E. Lerner
J. Letiche
F. Levy
S. Lewis
H. Liebhafsky
J. Lintner
T. C. Liu
M. Lovell
B. Malkiel
A. Manne
S. Marglin
J. Margolis
A. Marty
E. Mason
B. Massell
W. Massy
T. Mayer
D. McFadden
R. McKean
L. McKenzie
R. McKinnon
A. Meltzer
J. Melvin
J. Meyer
P. Meyer
M. Michaely
H. Miller
M. Miller
E. Mills
E. Mishan
F. Modigliani
J. Mooney
G. Morrison
L. Moses
M. Moyer
R. Musgrave
R. Muth
M. Nerlove
D. Netzer
J. Noren
D. North
W. Nutter
W. Oi
D. Orr
D. Paden
P. Pashigian
D. Patinkin
M. Pauly
J. Pechman
G. Perry
R. Pfouts
E. Phelps
A. Phillips
G. Pierson
J. Polak
J. Porterfield
R. Powell
J. Quirk
G. Ranis
L. Rapping
M. Reder
A. Rees
M. Reid
S. Reiter
R. Rhomberg
L. Ritter
A. Rivlin
E. Rolph
G. Rosenbluth
J. Rosse
T. Rothenberg
W. Salant
A. Sametz
P. Samuelson
T. Schultz
A. Scitovsky
R. Selden
W. Sharpe
E. Shaw
J. Sheahan
K. Shell
W. Shepherd
R. Sherman
C. Shoup
I. Silberman
V. Smith
D. Snider
R. Solow
H. Somers
J. Spengler
B. Stein
J. Stein
G. Stigler
K. Strand
D. Suits
R. Summers
P. Tangri
M. Taussig
R. Teigen
L. Telser
E. Thompson
G. Thompson
L. Thurow
J. Tobin

D. Tucker
J. Vanek
W. Vickrey
B. Ward
B. Weisbrod
L. Weiss
P. Wells
T. Whitin
O. Williamson
D. Winch
S. Winter
C. Wolf
P. Wonnacott
M. Yaari
L. Yeager
D. Yett

JOHN G. GURLEY, *Managing Editor*

JOURNAL OF ECONOMIC ABSTRACTS

REPORT OF THE EDITOR FOR THE YEAR ENDING DECEMBER 31, 1968

The transition to the *Journal of Economic Literature* has now been made. The *Journal of Economic Abstracts* has therefore ceased to exist in its earlier incarnation. The sole remaining function of the Editor is to report on expenditures for the financial year, December 1, 1967, to November 30, 1968. They were as follows:

Editor's salary	$1,500.00
Assistant editor's salary	1,000.00
Secretarial	500.00
Postage, telephone, supplies	48.14
Replacement of equipment	357.77
	$3,405.91

ARTHUR SMITHIES, *Editor*

REPORT OF COMMITTEE ON ECONOMIC EDUCATION

The primary goal of this Committee during 1967 and 1968 has been to stimulate experimentation with different means of teaching economics, followed by careful evaluation and measurement of results. Stress is being placed on the development of evaluated experiments so the results can be meaningfully made available to other institutions. We are convinced that this is one important way in which the profession can make further progress toward improving the efficiency of teaching economics.

We can report two completed projects during 1968:

1. The special task force charged with the development of new "Tests of Understanding in College Economics" has completed its work, and four test forms are currently available from the Psychological Corporation, 304 East 45th Street, New York, New York, 10017. There are two forms, statistically equivalent, developed for use in courses emphasizing elementary microeconomics, and two forms for courses emphasizing elementary macroeconomics. The primary purpose of these tests is to help individual institutions assess the results of their teaching and to permit cross-comparison of results with different teaching approaches at different institutions. All the individual questions and the tests as wholes have been carefully pretested and national norms have been established on the basis of a large nationwide sample. The Committee hopes that the profession will make wide use of these tests, and particularly that performance records on the standardized tests will be made available to this Committee whenever they are used. We expect to continue intensive work on the evaluation of alternative teaching methods, making use of these tests.

The TUCE task force included: Rendigs Fels, Chairman, G. L. Bach, William Bowen, R. A. Gordon, B. F. Haley, Paul Samuelson, and George Stigler from the profession; and Paul Dressel (Executive Secretary) and John Stalnaker as advisers from the field of educational psychology.

2. A nationwide analysis of the effectiveness of programmed learning in elementary economics, commissioned by the Joint Council on Economic Education at the suggestion of this Committee, has been completed by Professors Richard Attiyeh, G. L. Bach, and Keith Lumsden. This experiment has included over 4,000 students at approximately fifty schools of various types. A summary of the results is included in this volume, and more detailed results will be published elsewhere.

In addition, the Committee, working with the Joint Council on Economic Education, has in process the following enterprises:

a) Working with the cooperation of this Committee, the Joint Council on Economic Education has substantially completed plans for the inauguration of a *Journal on Economic Education*. This *Journal* is intended to improve the information flow on experiments and alternative approaches to the teaching of economics, primarily at the college level. Articles are expected to cover teaching experiments, evaluated results, and other matters of particular interest to those teaching college-level and senior high school economics. Barring unforeseen problems, the first issue of the *Journal* will appear in late autumn 1969. Establishment of the *Journal* is being financed by the Joint Council on Economic Education, with special funds from the Kazanjian Foundation. At the conclusion of two years, either the *Journal* will have to be self-supporting or additiona funds will be sought by the Joint Council. Thi Committee and the J.C.E.E. have prevailed or Henry H. Villard to serve as editor of the new *Journal*, and John Maher of the J.C.E.E. as managing editor. The original editorial board of the *Journal* will consist of members of this Committee plus distinguished educational psychologists. The intention is to publish the *Journal* twice each year; copy is in hand for the first two issues. We expect that this *Journal* may provide an outlet for reports on the teaching of economics which would not be appropriate for publication in the usual professional economic journals but which nonetheless merit dissemination and discussion. Th goal is to extend substantially the half-life o worthwhile experiments in the teaching of economics, on which there is presently no effective means of disseminating information.

b) Under J.C.E.E. sponsorship, the University of Michigan is conducting a special analysis of the effectiveness of alternative teaching methods, including various uses of teaching assistants, in the elementary course. This project includes the development of new measurement instruments, and analysis of the effects of attitudes and attitudinal changes on student learning in economics.

c) This Committee and the J.C.E.E. are planning a study of substantial size on the "lasting effects" of college elementary economics courses. Preliminary work on this project is being done by Professor Phillip Saunders.

d) The Joint Council on Economic Education, at the suggestion of this Committee, is planning to subsidize pilot-plant experimental courses in elementary economics at from three to six different institutions of differing character. Each of these new courses will be carefully evaluated, using appropriate instruments, and will be written up for possible use by other institutions, especially junior colleges and smaller liberal arts colleges which lack adequate teaching personnel to develop such courses on their own. Pilot projects have been tentatively discussed with three schools, and funds for them are being sought by the J.C.E.E.

e) Over the past three years, the Committee has worked with the J.C.E.E. to encourage preparation of a series of small monographs on economic topics to supplement regular teaching materials in the high schools. Six of these monographs are now in press, aimed primarily at high school courses in history and the social studies outside of economics, since economics is taught at many schools only as part of their "social studies" courses. Four others are in process, and two additional monographs may be commissioned in the near future. Each monograph has been prepared by a well-known economist, working with the assistance of a high school teacher or supervisor. Each is being thoroughly classroom tested before being released for publication.

The Visiting Scientists Program in Economics, for which the Association has received grants from the National Science Foundation over the past three years, has flourished. Professor Phillip Saunders is administering the program for the Committee. His report on the activities under the Visiting Scientists Program during the past year is printed below.

In conclusion, the Committee wishes to record a special vote of appreciation to Ben Lewis, who this year ended many years of service on this Committee, for his extraordinary contribution to the cause of better teaching in economics. For over a decade, he has been "Mr. Economic Education" to economists and to educators in the secondary schools. No one has done more to further humane and effective economics teaching. The profession owes him a large debt of gratitude.

G. L. BACH, *Chairman*

REPORT OF REPRESENTATIVE TO THE NATIONAL BUREAU OF ECONOMIC RESEARCH

Staff and Conference Programs. During 1968, staff research at the National Bureau of Economic Research continued in the following areas: economic growth—productivity, employment, and the behavior of prices and wages; human resources—education, health, and environmental factors; national income, consumption, and capital formation; business cycles and forecasting; financial institutions and processes; and international economic relations.

Among new and developing studies undertaken in 1968 were those on urban economics including industry location, transport planning and patterns of adaptation of housing stock by John R. Meyer, John F. Kain, H. James Brown, Royce Ginn, Mahlon R. Straszheim, and Raymond Struyk. An advisory committee with members from business, government, and universities was appointed to aid in the guidance of the urban economic studies.

Other new studies include the forecasting characteristics of econometric models by Yoel Haitovsky and Michael Evans; economic growth models by David T. Kresge; simulation of tax reforms by John Bossons; agricultural productivity by Franklin M. Fisher and Peter Temin; capital formation estimates by Robert J. Gordon; term structure of interest rates by Thomas J. Sargent; industrial diversification and integration by Michael Gort; patterns of managerial compensation by Wilbur G. Lewellen; and law and economics by Isaac Ehrlich.

Several regional economic studies are being developed, including those on unemployment patterns by David Gordon, migration patterns by Joseph Persky, and education by Dave O'Neill.

At the exploratory stage, Nancy and Richard Ruggles and others are investigating the possibility of establishing an economic research utility involving a large data bank and time-shared computer facilities. Alfred H. Conrad and Donald S. Shoup are involved in a cooperative effort with foreign economic research institutes to study the rates of technological diffusion in respective countries.

Research Fellows appointed for 1968–69 are William M. Landes, University of Chicago, Seiji Naya, University of Wisconsin, and Neil Wallace, University of Minnesota. David Levhari served as a Research Fellow in the summer of 1968.

Two conferences were sponsored by the Universities-National Bureau Committee for Economic Research: The Economics of Public Output (April, 1968) and Technology and Competition in International Trade (October, 1968). The Conference on Research in Income and Wealth sponsored a meeting on Education and Income (November, 1968). Conferences being planned are Income and Wealth conferences on Econometric Models of Cyclical Behavior (fall, 1969) and on International Comparisons of Real Product (spring, 1970), and Universities-National Bureau Committee conferences on International Mobility of Capital (winter 1969–70) and on Education as an Industry (fall, 1970).

Publications. Twelve National Bureau publications were issued in 1968 and three others were in press in December. Of these fifteen publications, three were volumes of conference proceedings, and the others were staff reports comprised of nine books, two occasional papers, and one technical paper.

Publication of a newsletter, *National Bureau Report,* was instituted in 1968. The *Report* briefly covers current activities of the Board and staff, conferences, new studies, recent and forthcoming publications, and lists manuscripts, papers, and reprints that are available from the National Bureau. Somewhat longer reports on research by members of the staff are published as supplements to the *Report.* During 1968 the *Report* was issued in January, May, and December with the following respective supplements: *Composite Indexes of Leading, Coinciding, and Lagging Indicators, 1948–67,* by Julius Shiskin and Geoffrey H. Moore; *Tariff Preferences for Less Developed Countries,* by Hal B. Lary;

and *The Growing Demand for Medical Care*, by Victor R. Fuchs.

Meetings of the Board of Directors in 1968. At the 49th annual meeting of the National Bureau, held March 18 and April 20, 1968, the Board adopted amendments to the bylaws which provided for some realignment of duties among National Bureau officers, for a change in operating basis from a calendar year to a fiscal year ending June 30, and for subsequent annual meetings on September 30. The first annual meeting on the new basis was also held in 1968.

During the year several members of the Board who have long served the Bureau with distinction and dedication retired and became Directors Emeriti: Harry W. Laidler, Joseph H. Willits, Harold M. Groves, and Gottfried Haberler.

A. G. Abramson, who had been elected a Director upon nomination of the National Association of Business Economists, died on October 6. George Cline Smith was elected as his successor.

Other newly elected representative Directors were Moses Abramovitz, Stanford University, Emilio G. Collado, Committee for Economic Development, Otto Eckstein, Harvard University, and Robert J. Lampman, University of Wisconsin. Newly elected Directors at Large were Robert A. Charpie, President, Bell & Howell Company; Eugene P. Foley, President, International Ore and Fertilizer Corporation; Eli Goldston, President, Eastern Gas and Fuel Associates; Vivian W. Henderson, President, Clark College; and J. Irwin Miller, Chairman of the Board, Cummins Engine Company. Theodore O. Yntema, who formerly represented the Committee for Economic Development, was also elected a Director at Large.

At the fall meeting a new position of Honorary Chairman of the Board was created, and Arthur F. Burns was elected to this post for an indefinite term. Theodore O. Yntema succeeded Mr. Burns as Chairman of the Board. Other officers elected were Walter W. Heller, Vice Chairman, John R. Meyer, President, Donald B. Woodward, Treasurer, Geoffrey H. Moore and Victor R. Fuchs, Vice Presidents-Research, Douglas H. Eldridge, Vice President-Administration, and Hal B. Lary, Director of International Studies.

The National Bureau's 48th Annual Report, "Toward Improved Social and Economic Measurement," was presented to the Board at the spring meeting. Copies of the report are available from the National Bureau upon request.

WILLARD L. THORP

REPORT OF REPRESENTATIVE TO THE INTERNATIONAL ECONOMIC ASSOCIATION

In 1968, the International Economic Association held its Congress, an affair scheduled every six years. There are usually several conferences held every year, attended only upon special invitation. The congresses held every six years are open to all members of the constituent national associations.

The Congress was held in Montreal, September 2–7. Its theme was "The Future of International Economic Relations." Papers were presented by H. Myint (Burma), H. G. Johnson (Canada), T. S. Khachaturov (U.S.S.R), A. Nove (U.K.), I. Vajda (Hungary), G. Ohlin (Sweden), A. Marchal, deceased (France), and T. Scitovsky (U.S.A.). Round table discussions were chaired by S. G. Triantis (Canada), P. N. Rasmussen (Denmark), M. Kaser (U.K.), and F. Machlup (U.S.A.). The presidential address was given by P. A. Samuelson (U.S.A.).

The meeting of the Council was attended by forty members, representing thirty-seven countries. The Executive Committee, consisting of twelve members, met and submitted to the Council nominations for a term of three years. E. Lundberg (Sweden) was elected President, J. Marchal (France), Vice-President, D. Patinkin (Israel), Treasurer, and E. A. G. Robinson (U.K.), Editor. Other members of the Executive Committee are : H. C. Eastman (Canada), J. Hicks (U.K.), N. Islam (Pakistan), T. S. Khachaturov (U.S.S.R.), J. H. G. Olivera (Brazil), P. N. Rasmussen (Denmark), P. A. Samuelson (U.S.A.), and O. Sik (Czechoslovakia).

Plans for the future include a regional conference on "Comparative Planning Experience in South Asia" (May, 1969); conference on "Mutual Economic Relations between North American and Europe" (September, 1969); conference on "The Essence of a Growth Model" (March, 1970); conference on "Market Relationships and Central Planning" (May, 1970); training workshops in Europe (summer 1970); and a refresher course in Pakistan (1969 or 1970).

Two volumes of proceedings of past conferences will appear soon, and a third has been dispatched to the printers. Four others are in preparation.

FRITZ MACHLUP

ECONOMICS INSTITUTE POLICY AND ADVISORY BOARD

This was a very eventful year in the life of the Economics Institute. First of all, the American Economic Association accepted a grant of $350,000 on behalf of the Institute. The grant, which is to apply to the five-year period beginning February 1, 1968, provides $204,000 with which to establish a revolving underwriting fund and $146,000 to be earmarked for fellowship aid to participants. The new grant supplements a balance of $34,591.29 remaining in the preceding grant which has been added to the revolving underwriting fund. In the opinion of the Policy and Advisory Board this puts the Institute in a sound financial position for the foreseeable future.

It should also be mentioned that the relations between the American Economic Association and the Economics Institute have been very much strengthened during the past year. The Association is the national grantee and the Policy and Advisory Board provides professional supervision. Furthermore, the University of Colorado has recognized the Institute as an autonomous academic unit under its Office of the Vice-President for Academic Affairs.

As in the past, the Institute continued to provide very valuable services to foreign students. During its first year of operation, in 1958, 36 students attended the Institute from 19 countries. In 1968, the total enrollment was 104 students from 31 different countries. These students are now studying at 39 universities and colleges in the United States. As in the past, the Institute has managed to attract a distinguished faculty and the level of instruction—in English, mathematics, and economics—appeared to be of very fine quality.

The current membership of the Policy and Advisory Board consists of Professors Neil Chamberlain (Columbia University), Gregg Lewis (University of Chicago), Douglass North (University of Washington), Vernon Ruttan (University of Minnesota), Joseph Spengler (Duke University), Henry Rosovsky (Harvard University), Chairman. Messrs. Chamberlain and Lewis will go off the Board by June of 1969 and their replacements should be nominated by the Executive Committee at its December meeting.

I would like to take this opportunity to comment on two further items. At a recent meeting of the Policy and Advisory Board, which was also attended by Professor Harold F. Williamson, we discussed the future of the American Economic Association's screening project. The Board felt that the screening project was of great potential value and that there might be considerable economies in combining it with the activities of the Economics Institute. Professor Owen shares this view. At the same time the Board also felt that adequate screening procedures for foreign students could not and should not be a shoestring operation. It would require additions to the Institute staff, perhaps in the form of a full-time person whose activities were entirely devoted to screening. And of course this costs money, which the Institute at present does not have.

HENRY ROSOVSKY

REPORT OF THE VISITING SCIENTIST PROGRAM IN ECONOMICS

The academic year 1967–68 marked the second year of the American Economic Association's participation in the Visiting Scientists Program sponsored by the National Science Foundation. The purposes of the program are to stimulate interest in economics among undergraduates in colleges and universities through visits by leaders in the profession and to provide opportunities for economics staff members and administrators at the visited institutions to discuss their teaching and research problems with the visiting economist. Emphasis is placed on visits to smaller colleges and universities whose major focus in economics is at the undergraduate level.

As in the preceeding year, the 1967–68 program was administered by Professor Phillip Saunders, of Carnegie-Mellon University, under the direction of the Association's Committee on Economic Education, Professor G. L. Bach, of Stanford University, Chairman. All financial transactions were executed and recorded through the offices of the A.E.A.'s Secretary-Treasurer, Professor Harold Williamson.

The summary figures given below show that the number of schools visited and the total number of visit days increased during the second year of the program, and that an increase in cash contributions by the schools visited and a reduction in the official NSF honorarium resulted in a lower total cost in 1967–68 than in 1966–67.

The 1967–68 visits covered 22 different states from Massachusetts to Oregon and from North Dakota to Mississippi. In arranging visits to the schools selected, the informal roster of distinguished economists established in the preceeding year was used extensively, and it was supplemented with direct arrangements in a few specific cases. The continuing cooperation and enthusiasm of the individual economists asked to make visits is one of the most heartening aspects of the program.

The past success of the Visiting Scientists Program in Economics led the National Science Foundation to extend the program for a third year with an increased budget for academic year 1968–69. The A.E.A.'s Committee on Economic Education then met in New York on September 11, 1968, and decided to employ essentially the same criteria for selecting schools for visits in 1968–69 as had been used in the preceding two years. These criteria provided that the program director, in consultation with the Chairman and other members of the Committee as appropriate, would give first preference to: schools with no extensive graduate programs, certainly not beyond the master's level; schools geographically located in areas where they have no easy access to leading economists; schools whose applications showed some imagination and thought as to how the visit would be effectively used; schools that were neither already of high excellence or had very low standards, on the assumption that visits to "in-between" schools would be most useful.

It was also decided that information describing the program should be mailed to both public and private predominantly Negro colleges not covered by the United Negro College Fund.

To date, 46 visits have been tentatively arranged under the 1968–69 program. Since most of these visits are scheduled for the spring term, however, only fragmentary results are available concerning the third-year program. Nevertheless, there is no evidence to indicate that the program will be any less successful than it was in preceding years.

PHILLIP SAUNDERS

School Year	Number of Schools Visited	Number of Visit Days	Expenditure	Cash Contribution	Cost to NSF
1966–67	20	34	$4,903	$223	$4,680
1967–68	32	43	5,274	684	4,590

REPORT OF THE CENSUS ADVISORY COMMITTEE

The Census Advisory Committee of the American Economic Association met with the Director and Staff of the Bureau of the Census on January 19, 1968. Among the more important subjects discussed at the meeting were: proposed legislation introduced in Congress to limit the number of questions to be asked on the Decennial Censuses of Population and Housing; expanded content of the Bureau's *Business Cycle Developments* report to make it into a general-purpose economic indicators report; evaluation of the benefits of he Census Bureau's output in compliance with the Budget Bureau's Planning-Programming-Budgeting System; means of making the detailed data collected more accessible and more useful for a wide variety of uses and sers; and improvement in the Bureau's statistical programs in the fields of the value of new construction put-in-place and the service trades.

The Advisory Committee's Technical Subcommittee on Business Cycle Developments met with the Director and Staff of the Bureau of the Census on May 13, 1968. The main business of the Subcommittee meeting was to discuss proposals for broadening and otherwise improving the valuable Census monthly publication, *Business Cycle Developments.* Many of the suggestions made are already embodied in *Business Conditions Digest,* as the publication was entitled beginning with the November, 1968, issue. Members of the Association are urged to acquaint themselves with the new "BCD."

SOLOMON FABRICANT, *Chairman*

REPORT OF REPRESENTATIVE TO THE NATIONAL RESEARCH COUNCIL, DIVISION OF BEHAVIORAL SCIENCES

The American Economic Association appoints three members to the Behavioral Science Division of the National Research Council—National Academy of Sciences. Currently those three are Harold Barnett, Lionel McKenzie, and Guy Orcutt. One of the three serves on the Executive Committee, currently Harold Barnett. In addition there are members at large appointed both to the Executive Committee and the Division (currently Guy Orcutt and James Morgan) or just to the Behavioral Science Division (currently D. Gale Johnson and Joseph J. Spengler). Chairman of the Division for two years ending June, 1970, is Herbert Simon. For the two years after that the Chairman will be James Morgan.

Since the main function of the Behavioral Science Division is to establish *ad hoc* advisory committees to advise various government offices and bureaus and departments or to deliberate on the needs of the behavioral and social sciences more broadly, a brief list of current committees follows: The largest effort is the Behavioral and Social Sciences Survey Committees where Carl Kaysen and Robert Solow serve on the Central Planning Committee, and jointly chair the Economics Panel. The other members of the Economics Panel are: Hollis Chenery, Arnold Harberger, Dale Jorgenson, James Morgan, Guy Orcutt, Joseph Pechman, Melvin Reder, Henri Theil, and James Tobin.

Papers by the members of the Economics Panel on the accomplishments and future needs of various areas of economics may be published separately, but they also feed into the total report.

The Environmental Studies Board is jointly sponsored by the National Academy of Engineering and the National Academy of Sciences and deals with problems of the quality of the environment, including pollution. Harvey Perloff is a member.

The Committee on Information in the Behavioral Sciences continues, chaired by David Easton (political science), with Richard Ruggles serving on the Committee and on its Subcommittee on Structure and Social Organization of Information Systems. Fritz Machlup also serves on the Subcommittee on Needs and Dissemination Practices.

The Advisory Committee on Government Programs in the Behavioral Sciences on which Thomas Schelling and Joseph Spengler serve, issued its report in September, 1968, under the title, *The Behavioral Sciences and the Federal Government* (NAS Publication 1680).

The Committee on International Relations in the Behavioral Sciences was established in 1967, chaired by sociologist Seymour M. Lipset. Lawrence Klein serves on that committee. It advises the Social Science Subcommittee on the U.S. National Commission for UNESCO, and has developed a study of nongovernmental international organizations and activities in the behavioral and social sciences.

In the spring of 1968 a new committee was formed, originally called the Committee on Biological Research, Social Behavior, and Social Policy, but later shortened to the Committee on the Life Sciences and Social Policy. It is concerned with the issues arising out of modern biological advances (transplants, genetic alteration, etc.)

Two committees were set up to advise the Department of Housing and Urban Development—one on Social and Behavioral Urban Research, chaired by Raymond Bauer, and involving Harvey Brazer and Abel Wolman as members.

An Advisory Committee to the Arms Control and Disarmament Agency on Support of Dissertation Research in the Social and Behavioral Sciences was set up in June, 1968, with Abram Bergson as one of its members.

A Committee on Basic Research in Education, to advise the U.S. Office of Education, was established in June, 1968, with Fritz Machlup as one of its members.

In a more informal vein, the Division consulted with Philip Handler, Chairman of the National Science Board, in designing the call for a Commission on the Social Sciences, advisory to the National Science Foundation. It is expected to report this winter. Paul Samuelson serves on that Commission, which by the NSF charter must contain as many members outside the sciences involved as inside.

A recurring problem to economists serving on the Behavioral Sciences Division is knowing who in the profession is interested in and capable of serving on the various committees. Any of the present members of the Division would welcome suggestions both for people and for problem areas in which the Division should initiate studies.

James N. Morgan

PUBLICATIONS
OF THE
AMERICAN ECONOMIC ASSOCIATION

1969

FIRST SERIES

Volume I, 1886

1. Report of Organization of the American Economic Association. Pp. 46.
2-3. Relation of the Modern Municipality to the Gas Supply. By E. J. James. Pp. 66.
4. Co-operation in a Western City. By Albert Shaw. Pp. 106.
5. Co-operation in New England. By E. W. Bemis. Pp. 136.
6. Relation of the State to Industrial Action. By H. C. Adams. Pp. 85.

Volume II, 1887

1. Three Phases of Co-operation in the West. By Amos G. Warner. Pp. 119.
2. Historical Sketch of the Finance of Pennsylvania. By T. K. Worthington. Pp. 106.
3. The Railway Question. By Edmund J. James. Pp. 68.
4. Early History of the English Woolen Industry. By W. J. Ashley. Pp. 85.
5. Mediaeval Guilds of England. By E. R. A. Seligman. Pp. 113.
6. Relation of Modern Municipalities to Quasi-public Works. By H. C. Adams and others. Pp. 87.

Volume III, 1888

1. Statistics in College, by C. D. Wright; Sociology and Political Economy, by F. H. Giddings; The Legal-Tender Decisions, by E. J. James. Pp. 80.
2. Capital and Its Earnings. By John B. Clark. Pp. 69.
3. The Manual Laboring Class, by F. A. Walker; Mine Labor in the Hocking Valley, by E. W. Bemis; Report of the Second Annual Meeting. Pp. 86.
4-5. Statistics and Economics. By Richmond Mayo-Smith. Pp. 127.
6. The Stability of Prices. By Simon N. Patten. Pp. 64.

Volume IV, 1889

1. Contributions to the Wages Question: The Theory of Wages, by Stuart Wood; Possibility of a Scientific Law of Wages, by J. B. Clark. Pp. 69.
2. Socialism in England. By Sidney Webb. Pp. 73.
3. Road Legislation for the American State. By J. W. Jenks. Pp. 83.
4. Third Annual Meeting: Report of the Proceedings. Pp. 123.
5. Malthus and Ricardo, by S. N. Patten; The Study of Statistics, by D. R. Dewey; Analysis in Political Economy, by W. W. Folwell. Pp. 69.
6. An Honest Dollar. By E. Benjamin Andrews. Pp. 50.

Volume V, 1890

1. The Industrial Transition in Japan. By Yeijro Ono. Pp. 122.
2. Two Essays on Child-Labor. By W. F. Willoughby and Clare de Graffenried. Pp. 150.
3-4. Papers on the Canal Question. By E. J. James and L. M. Haupt. Pp. 85.
5. History of the New York Property Tax. By J. C. Schwab. Pp. 108.
6. Educational Value of Political Economy. By S. N. Patten. Pp. 36.

Volume VI, 1891

1-2. Fourth Annual Meeting: Reports, Papers, Discussions.
3. Government Forestry. Papers by Pinchot, Bowers, and Fernow. Pp. 102.
4-5. Municipal Ownership of Gas in the U.S. By E. W. Bemis. Pp. 185.
6. State Railroad Commissions. By F. C. Clark. Pp. 110.

Volume VII, 1892

1. The Silver Situation in the United States. By F. W. Taussig. Pp. 118.
2-3. Shifting and Incidence of Taxation. By E. R. A. Seligman. Pp. 424. (Revised.)
4-5. Sinking Funds. By Edward A. Ross. Pp. 106.
6. The Reciprocity Treaty with Canada of 1854. By F. E. Haynes. Pp. 70.

Volume VIII, 1893

1. Fifth Annual Meeting: Report of the Proceeding. Pp. 130.
2-3. Housing of the Poor in American Cities. By M. T. Reynolds. Pp. 132.
4-5. Public Assistance of the Poor in France. By E. G. Balch. Pp. 180.
6. First Stages of the Tariff Policy of the U.S. By William Hill. Pp. 162.

Volume IX, 1894

Sixth Annual Meeting: Handbook and Report. Pp. 73.
1-2. Progressive Taxation in Theory and Practice. By Edwin R. A. Seligman. PP. 222. *(See 1908, No. 4.)*
3. The Theory of Transportation. By C. H. Cooley. Pp. 148.
4. Sir William Petty. By Wilson Lloyd Bevan. Pp. 102.
5-6. Papers on Labor Problems. By J. B. Clark, C. D. Wright, D. R. Dewey, A. T. Hadley, and J. G. Brooks. Pp. 94.

Volume X, 1895

Seventh Annual Meeting: Handbook and Report. Pp. 183.
1-3. The Canadian Banking System, 1817-1890. By R. M. Breckenridge. Pp. 478.
4. Poor Laws of Massachusetts and New York. By John Cummings. Pp. 136.
5-6. Letters of Ricardo to McCulloch, 1816-1823. Edited by J. H. Hollander. Pp. 204.

Volume XI, 1896

1-3. Race Traits and Tendencies of the American Negro. By F. L. Hoffman. Pp. 330.
4. Appreciation and Interest. By Irving Fisher. Pp. 110.
General Index to Volumes I-XI (1886-1896).

ECONOMIC STUDIES

($2.50 per volume)

Volume I, 1896

Eighth Annual Meeting: Handbook and Report. Pp. 78.
1. The Theory of Economic Progress, by J. B. Clark; The Relation of Changes in the Volume of the Currency to Prosperity, by F. A. Walker. Pp. 46.
2. The Adjustment of Wages to Efficiency. Three Papers: Gain Sharing, by H. R. Towne; The Premium Plan, by F. A. Halsey; A Piece-Rate System, by F. W. Taylor. Pp. 83.
3. The Populist Movement. By Frank L. McVey. Pp. 81.
4. The Present Monetary Situation. By W. Lexis; translated by John Cummings. Pp. 72.
5-6 The Street Railway Problem in Cleveland. By W. R. Hopkins. Pp. 94.

Volume II, 1897

Ninth Annual Meeting: Handbook and Report. Pp. 162.
1. Economics and Jurisprudence. By Henry C. Adams. Pp. 48.
2. The Saloon Question in Chicago. By John E. George. Pp. 62.
3. The General Property Tax in California. By C. C. Plehn. Pp. 88.
4. Area and Population of the United States at the Eleventh Census. By W. F. Willcox. Pp. 60.
5. A Discussion Concerning the Currencies of the British Plantations in America, etc. By William Douglass. Edited by C. J. Bullock. Pp. 228.
6. Density and Distribution of Population in the United States at the Eleventh Census. By W. F. Willcox. Pp. 79.

Volume III, 1898

Tenth Annual Meeting: Handbook and Report. Pp. 136.
1. Government by Injunction. By William H. Dunbar. Pp. 44.
2. Economic Aspects of Railroad Receiverships. By H. H. Swain. Pp. 118.
3. The Ohio Tax Inquisitor Law. By T. N. Carver. Pp. 50.
4. The American Federation of Labor. By Morton A. Aldrich. Pp. 54.
5. Housing of the Working People in Yonkers. By E. L. Bogart. Pp. 82.
6. The State Purchase of Railways in Switzerland. By Horace Michelie; translated by John Cummings. Pp. 72.

Volume IV, 1899

Eleventh Annual Meeting: Handbook and Report. Pp. 126.
1. I. Economics and Politics. By A. T. Hadley. II. Report on Currency Reform. III. Report on the Twelfth Census. Pp. 70.
2. Personal Competition. By Charles H. Cooley. Pp. 104.
3. Economics as a School Study. By F. R. Clow. Pp. 72.

4-5. The English Income Tax. By J. A. Hill. Pp. 162.

6. Effects of Recent Changes in Monetary Standards upon the Distribution of Wealth. By F. S. Kinder. Pp. 91.

NEW SERIES

1. The Cotton Industry. By M. B. Hammond. Pp. 382.
2. Scope and Method of the Twelfth Census. Critical discussion by over twenty statistical experts. Pp. 625.

THIRD SERIES

Note—During 1896-1899 *the Association issued its publications in two series, viz., the bimonthly* Economic Studies, *and the "New Series" of larger monographs printed at irregular intervals. In* 1900 *it reverted to the policy of issuing its monographs, now called the "Third Series" of the publications, at quarterly intervals.*

Price per volume, $4.00.

Volume I, 1900

1. Twelfth Annual Meeting: Papers on Economic Theory and Political Morality; Trusts; Railroad Problems; Public Finance; Consumers' League; Twelfth Census. Pp. 186.
2. The End of Villeinage in England. By T. W. Page. Pp. 99.
3. Essays in Colonial Finance. By Members of the Association. Pp. 303.
4. Currency and Banking in the Province of Massachusetts Bay. By A. McF. Davis. Part I: Currency. Pp. 464 + 91 photogravure plates.

Volume II, 1901

1. Thirteenth Annual Meeting: Papers on Competition; Commercial Education; Economic Theory; Taxation of Quasi-public Corporations; Municipal Accounts. Pp. 300.
2. Currency and Banking. By A. McF. Davis. Part II: Banking. Pp. 341 + 18 photogravure plates.
3. Theory of Value before Adam Smith. By Hannah R. Sewall. Pp. 132.
4. Administration of City Finances in the U.S. By F. R. Clow. Pp. 144.

Volume III, 1902

1. Fourteenth Annual Meeting: Papers on International Trade; Industrial Policy; Public Finance; Protective Tariff; Negro Problem; Arbitration of Labor Disputes in Coal Mining Industry; Porto Rican Finance; Economic History. Pp. 400.
2. The Negro in Africa and America. By Joseph A. Tillinghast. Pp. 240.
3. Taxation in New Hampshire. By M. H. Robinson. Pp. 232.
4. Rent in Modern Economic Theory. By Alvin S. Johnson. Pp. 136.

Volume IV, 1903

1. Fifteenth Annual Meeting: Papers and Discussions on Economic and Social Progress; Trade Unions and the Open Shop; Railway Regulation; Theory of Wages; Theory of Rent; Oriental Currency Problem. Pp. 298.
2. Ethnic Factors in the Population of Boston. By F. A. Bushee. Pp. 171.
3. History of Contract Labor in the Hawaiian Islands. By Catharine Coman. Pp. 74.
4. The Income Tax in the Commonwealth of the United States. By Delos O. Kinsman. Pp. 134.

Volume V, 1904

Sixteenth Annual Meeting. Papers and Proceedings published in two parts.

1. Part I—Papers and Discussions on Southern Agricultural and Industrial Problems; Social Aspects of Economic Law; Relations between Rent and Interest. Pp. 240.
2. Part II—Papers and Discussions on The Management of the Surplus Reserve; Theory of Loan Credit in Relation to Corporation Economics; State Taxation of Interstate Commerce; Trusts; Theory of Social Causation. Pp. 203.
3. Monopolistic Combinations in the German Coal Industry. By Francis Walker. Pp. 340.
4. The Influence of Farm Machinery on Production and Labor. By Hadley Winfield Quaintance. Pp. 110.

Volume VI, 1905

Seventeenth Annual Meeting. Papers and Proceedings published in two parts.

1. Part I—Papers and Discussions on the Doctrine of Free Trade; Theory of Prices; Theory of Money; Open Shop or Closed Shop. Pp. 226.
2. Part II—Papers and Discussions on Government Interference with Industrial Combinations; Regulation of Railway Rates; Taxation of Railways; Preferential Tariffs and Reciprocity; Inclosure Movement; Economic History of the United States. Pp. 270.
3. The History and Theory of Shipping Subsidies. By R. Meeker. Pp. 230.
4. Factory Legislation in the State of New York. By F. R. Fairchild. Pp. 218.

Volume VII, 1906

1. Eighteenth Annual Meeting: Papers and Discussions on The Love of Wealth and the Public Service; Theory of Distribution; Governmental Regulation of Railway Rates; Municipal Ownership; Labor Disputes; The Economic Future of the Negro. Pp. 325.
2. Railroad Rate Control. By H. S. Smalley. Pp. 147.
3. On Collective Phenomena and the Scientific Value of Statistical Data. By E. G. F. Gryzanovski. Pp. 48.

Handbook of the Association, 1906. Pp. 48.

4. The Taxation of the Gross Receipts of Railways in Wisconsin. By G. E. Snider. Pp. 138.

Volume VIII, 1907

1. Nineteenth Annual Meeting: Papers and Discussions on Modern Standards of Business Honor; Wages as Determined by Arbitration; Commercial Education; Money and Banking; Western Civilization and Birth Rate; Economic History; Government Regulation of Insurance; Trusts and Tariff; Child Labor. Pp. 268.
2. Historical Sketch of the Finances and Financial Policy of Massachusetts from 1780 to 1905. By C. J. Bullock. Pp. 144.

Handbook of the Association, 1907. Pp. 50.

3. The Labor Legislation of Connecticut. By Alba M. Edwards. Pp. 322.
4. The Growth of Large Fortunes. By G. P. Watkins. Pp. 170.

Volume IX, 1908

Handbook of the Association, 1908. Pp. 49.

1. Twentieth Annual Meeting: Papers and Discussions on Principles of Government Control of Business; Are Savings Income; Agricultural Economics; Money and Banking; Agreements in Political Economy; Labor Legislation; Relation of the Federal Treasury to the Money Market; Public Service Commissions. Pp. 311.
2. Chicago Traction. By R. E. Heilman. Pp. 131.
3. Factory Legislation of Rhode Island. By J. K. Towles. Pp. 119.
4. Progressive Taxation in Theory and Practice. Revised Edition. By E. R. A. Seligman. Pp. 334.

Volume X, 1909

1. Twenty-first Annual Meeting: Papers and Discussions on The Making of Economic Literature; Collective Bargaining; Round Table on Accounting; Labor Legislation; Employers' Liability; Canadian Industrial-Disputes Act; Modern Industry and Family Life; Agricultural Economics; Transportation; Revision of the Tariff; A Central Bank; The National Monetary Commission; Capitalization of Public Service Corporations in Massachusetts. Pp. 432.
2. Handbook of the Association, 1909. Pp. 59.
3. The Printers. By George E. Barnett. Pp. 379.
4. Life Insurance Reform in New York. By W. H. Price. Pp. 95.

Volume XI, 1910

1. Twenty-second Annual Meeting: Papers and Discussions on History of the Association; Observation in Economics; Economic Dynamics; Theory of Wages; Country Life; Valuation of Public Service Corporations; Trusts; Taxation. Pp. 386.
2. Handbook of the Association, 1910. Pp. 79.
3. The Child Labor Policy of New Jersey. By A. S. Field. Pp. 229.
4. The American Silk Industry and the Tariff. By F. R. Mason. Pp. 178.

THE ECONOMIC BULLETIN

Published quarterly in 1908, 1909, and 1910, and containing personal notes, news of the economic world, announcements of special lines of investigation, and a classified and annotated bibliography of the current books and articles on economic subjects.

Volume I, 1908. Four numbers.
Volume II, 1909. Four numbers.
Volume III, 1910. Four numbers.

FOURTH SERIES

Volume I, 1911

The American Economic Review, March, June, September, and December
Twenty-third Annual Meeting:
Papers and Discussions on the Significance of a Comprehensive System of Education; Money and Prices; The Ricardo Centenary; Accounting; Canals and Railways; Population and Immigration; Labor Legislation; Taxation; A Definition of Socialism; Competition in the Fur Trade. Pp. 388.
Handbook of the Association, 1911.

Volume II, 1912

The American Economic Review, March, June, September, and December.
Supplement.—Twenty-fourth Annual Meeting:
Papers and Proceedings. The Economic Utilization of History; Tariff Legislation; The Federal Budget; Rural Conditions; Selection of Population by Migration; The Price Concept; An International Commission on the Cost of Living; Industrial Efficiency. Pp. 146.

Volume III, 1913

The American Economic Review, March, June, September, and December
Supplement.—Twenty-fifth Annual Meeting:
Papers and Proceedings. Population or Prosperity; The Rising Cost of Living—Standardizing the Dollar; Banking Reform; Theories of Distribution; Farm Management; Governmental Price Regulation. Pp. 155.
Supplement.—Handbook of the Association, 1913.

Volume IV, 1914

The American Economic Review, March, June, September, and December
Supplement.—Twenty-sixth Annual Meeting:
Papers and Proceedings. The Increasing Governmental Control of Economic Life; The Control of Public Utilities; Railroad Rate Making; Syndicalism; Trust Decisions and Business. Pp. 211.

Volume V, 1915

The American Economic Review, March, June, September, and December
Supplement.—Twenty-seventh Annual Meeting:
Papers and Proceedings. Economics and the Law; Regulation of the Stock Exchanges; Market Distribution; Statistical Work of the United States Government; Relation of Education to Industrial Efficiency; The Effect of Income and Inheritance Taxes on the Distribution of Wealth; Public Regulation of Wages. Pp. 323.

Volume VI, 1916

The American Economic Review, March, June, September, and December
Supplement.—Twenty-eighth Annual Meeting:
Papers and Proceedings. The Apportionment of Representatives; Effect of the War on Foreign Trade; Budget Making and the Increased Cost of Government; Economic Costs of War; Economic Theorizing and Scientific Progress; The Role of Money in Economic Theory; Price Maintenance; Investment of Foreign Capital. Pp. 248.
Supplement.—Handbook of the Association, 1916.

Volume VII, 1917

The American Economic Review, March, June, September, and December
Supplement.—Twenty-ninth Annual Meeting:
Papers and Proceedings. The National Point of View in Economics; Landed Property; Two Dimensions of Economic Productivity; Some Social Surveys in Iowa; The Land Problem and Rural Welfare; The Federal Farm Loan Act; Statistics of the Concentration of Wealth; Gold Supply at the Close of the War; Loans and Taxes in War Finance; Problems of Population after the War; Some Phases of the Minimum Wage. Pp. 275.
Supplement.—Index to the Publications, 1886-1910.

Volume VIII, 1918

The American Economic Review, March, June, September, and December
Supplement.—Thirtieth Annual Meeting:
Papers and Proceedings. Economic Reconstruction; Federal Taxes upon Income and Excess Profits; Land Utilization and Colonization; Federal Valuation of Railroads; Co-ordination of Employment Bureaus; Control of the Acquisition of Wealth; Motives in Economic Life; Price-Fixing Problems of Governmental Efficiency; Economic Alliances and Tariff Adjustments. Pp. 317.

Volume IX, 1919

The American Economic Review, March, June, September, and December
Supplement.—Thirty-first Annual Meeting:
Papers and Proceedings. Economists in Public Service; Interest on Investment a Manufacturing Cost Factor; Control of Marketing Methods and Costs; War and the Supply of Capital; War and the Rate of Interest; Index of the Cost of Living; Securing the Initiative of the Workman; A Legal Dismissal Wage; After-War Gold Policies; Foreign Exchange; Stabilizing the Dollar; Tenancy of Landed Property; Price-Fixing; Economic

Theory in an Era of Readjustment; Psychology and Economics; The Open Door and Colonial Policy; Reports of Committees on Foreign Trade and Purchasing Power of Money. Pp. 368.

Supplement No. 2—Report of the Committee on War Finance. Pp. 142.

Supplement No. 3—Handbook of the Association, 1919.

Volume X, 1920

The American Economic Review, March, June, September, and December

Supplement.—Thirty-second Annual Meeting:

Papers and Proceedings. Excess Profits Taxation; Germany's Reparation Payments; International Supervision over Foreign Investments; Results of a Balanced Industrial System; Employee's Representation in Management of Industry; Prices and Reconstruction; Banking Policy and Prices; Large-Scale Marketing; Reports of Committees on Foreign Trade, Co-ordination in Taxation, Census Advisory Committee. Pp. 278.

Supplement No. 2—Taxation of Excess Profits in Great Britain. Pp. 244.

Volume XI, 1921

The American Economic Review, March, June, September, and December

Supplement.—Thirty-third Annual Meeting:

Papers and Proceedings. The Railroad Situation; Our Foreign Trade Balance; Problems of the Bituminous Coal Industry; Traditional Economic Theory; Non-Euclidean Economics; Federal Taxation of Profits and Income; Teaching of Elementary Economics. Pp. 194.

Supplement No. 2.—Karelsen Prize Essays, on What Can a Man Afford? Pp. 118.

Volume XII, 1922

The American Economic Review, March, June, September, and December

Supplement.—Thirty-fourth Annual Meeting:

Papers and Proceedings. Business Cycles; American Trade Unionism; The Railroads; Workmen's Compensation; Federation in Central America; Teaching of Elementary Economics; The Chain Store Grocer; Economics and Ethics. Pp. 194.

Supplement No. 2—Handbook of the Association, 1922.

Volume XIII, 1923

The American Economic Review, March, June, September, and December

Supplement.—Thirty-fifth Annual Meeting:

Papers and Proceedings. Trend of Prices; Trend of Rate of Interest and Investment; Outlook for Wages and Employment; Overhead Costs; Commercial Rent and Profits; Labor Turnover; Factors in Wage Determinations; Income of Farmers; Large-Scale Production and Merchandising; Marketing Farm Products; Bureau of Business Research. Pp. 293.

Volume XIV, 1924

The American Economic Review, March, June, September, and December

Supplement.—Thirty-sixth Annual Meeting:

Papers and Proceedings. International Trade and Commercial Policy; Railroad Consolidation; Economic Theory; Transportation; American Foreign Trade; Marketing. Pp. 192.

Supplement No. 2—Handbook of the Association, 1924.

Volume XV, 1925

The American Economic Review, March, June, September, and December

Supplement.—Thirty-seventh Annual Meeting:

Papers and Proceedings. The Economics of Advertising; Problems of Economic Theory; Transportation; Marketing; Giant Power; The Teaching of Business and Economics; Business Administration; Monetary Stabilization; Foreign Service Training; Highway Economics; Psychological Problems of Industry. Pp. 165.

Supplement No. 2.—Babson Prize Essay, on Forecasting the Price of Hogs. Pp. 22.

Volume XVI, 1926

The American Economic Review, March, June, September, and December

Supplement.—Thirty-eighth Annual Meeting:
Papers and Proceedings. Movement of Real Wages; Teaching of Economics; Consuming Power of Labor and Business Fluctuations; Economic Problems Involved in the Payment of International Debts; Economics and Geography; Agriculture in Our National Policy; Tariff Making; Trade Associations; Theory of Wages; Reducing the Costs of Marketing; Topics in Economic History; Railway Problems; Land Economics; Federal Reserve Policies. Pp. 353.
Supplement No. 2.—Handbook of the Association, 1926.

Volume XVII, 1927

The American Economic Review, March, June, September, and December
Supplement.—Thirty-ninth Annual Meeting:
Papers and Proceedings. Economics of Prohibition; Economic History; Use of the Quantitative Method in the Study of Economic Theory; Present-Day Corporation Problems; American Practices Analogous to Foreign Controls over Raw Materials; Marketing; Interest Theory and Price Movements; Problem of Effective Public Utility Regulation; Immigration Restriction—Economic Results and Prospects; Family Budgets; Motor Transportation in the United States. Pp. 218.
Supplement No. 2.—Report of the Dinner in Honor of Professor John Bates Clark. Pp. 18.

Volume XVIII, 1928

The American Economic Review, March, June, September, and December
Supplement.—Fortieth Annual Meeting:
Papers and Proceedings. Land Economics; Marketing; Present Status and Future Prospects of Quantitative Economics; Post-War Fluctuations of Commodity Prices; Relationship between Departments of Economics and Collegiate Schools of Business; Economic History; Simplification of the Federal Income Tax; Economic Significance of the Increased Efficiency of American Industry; An Approach to the Law of Production and Its Relation to the Welfare of the Wage-Earner; Meaning of Valuation; Railroad Valuation with Special Reference to the O'Fallon Decision; Interest Rates As Factors in the Business Cycle; Should the Debt Settlements Be Revised; An Examination of the Reasons for Revision of the Debt Settlements. Pp. 305.
Supplement No. 2.—Handbook of the Association, 1928.

Volume XIX, 1929

The American Economic Review, March, June, September, and December
Supplement.—Forty-first Annual Meeting:
Papers and Proceedings. Market Shifts, Price Movements, and Employment; Some Observations on Unemployment Insurance; Marketing; Land Economics; Law and Economics; Price Stabilization; London and the Trade Cycle; Federal Reserve Policy and Brokers' Loans; Central Planning of Production in Soviet Russia; International Differences in the Labor Movement; Tariff Making in the United States; Economic History; Locality Distribution of Industries; Regulation of Electric Light and Power Utilities; An Inductive Study of Publicly Owned and Operated vs. Privately Owned but Regulated Public Utilities; Regulation of the Common Carrier; Commercial Motor Vehicle and the Public. Pp. 284.

Volume XX, 1930

The American Economic Review, March, June, September, and December
Supplement.—Forty-second Annual Meeting:
Papers and Proceedings. Economic History; Public Works Plan and Unemployment; Theory of Economic Dynamics as Related to Industrial Instability; Chief Economic Problems of Mexico; Reparations Settlement and the International Flow of Capital; Federal Reserve Board—Its Problems and Policy; Economic and Social Consequences of Mechanization in Agriculture and Industry. Pp. 214.

Volume XXI, 1931

The American Economic Review, March, June, September, and December

Supplement.—Forty-third Annual Meeting:

Papers and Proceedings. Decline of Laissez Faire; Small Loan Business; Social and Economic Aspects of Chain Stores; Russian Economic Situation; Trustification and Economic Theory; Persistence of the Merger Movement; Program of the Federal Farm Board; Social Implications of Restriction of Agricultural Output; Land Economics and Real Estate; Institutionalism—What It Is and What It Hopes to Become; An Approach to World Economics; International Industrial Relations—Migration of Enterprise and Policies Affecting It; World-Wide Depression of 1930; Present Depression—A Tentative Diagnosis; Power and Propaganda; Failure of Electric Light and Power Regulation and Some Proposed Remedies. Pp. 302.

Supplement No. 2.—Handbook of the Association, 1931.

Volume XXII, 1932

The American Economic Review, March, June, September, and December

Supplement.—Forty-fourth Annual Meeting:

Papers and Proceedings. Private Enterprise in Economic History; Shorter Working Time and Unemployment; Quantitative Economics; Theory of Technological Progress and the Dislocation of Employment; Measurement of Productivity Changes and the Displacement of Labor; Stabilization of Business and Employment; Principle of Planning and the Institution of Laissez Faire; Institutional Economics; Elasticity of Demand as a Useful Marketing Concept; Investments of Life Insurance Companies; Real Estate in the Business Cycle; Investments and National Policy of the United States in Latin America; Recent Changes in the Character of Bank Liabilities and the Problem of Bank Reserves; Bank Failures in the United States; Transportation by Rail and Otherwise; Our Changing Transportation System. Pp. 306.

Volume XXIII, 1933

The American Economic Review, March, June, September, and December

Supplement.—Forty-fifth Annual Meeting:

Papers and Proceedings. Rise of Monopoly in the United States; Record of Insurance in the Depression; Some Theoretical Aspects of Unemployment Reserves; The Economics of Unemployment Relief; American Economic Thought; Formation of Capital; Measurement and Relation to Economic Instability; Size of Business Unit as a Factor in Efficiency of Marketing; Reserve Bank Policy and Economic Planning; Federal Reserve Policy in World Monetary Chaos; Tariff Reform: The Case for Bargaining; Speculation in Suburban Lands; Real Estate Speculation and the Depression. Pp. 206.

Supplement No. 2.—Handbook of the Association, 1933.

Volume XXIV, 1934

The American Economic Review, March, June, September, and December

Supplement.—Forty-sixth Annual Meeting:

Papers and Proceedings. The History of Recovery; Public Utilities in the Depression; Imperfect Competition; Fundamentals of a National Transportation Policy; Correlation of Rail and Highway Transportation; Marketing under Recovery Legislation; Economics of the Recovery Act; Measurement of Unemployment; Controlled Inflation; Banking Act of 1933—An Appraisal; Some Statistics on the Gold Situation; The Problem of Tax Delinquency; The Problem of Expanding Governmental Activities; The Economics of Public Works. Pp. 224.

Volume XXV, 1935

The American Economic Review, March, June, September, and December

Supplement.—Forty-seventh Annual Meeting:

Papers and Proceedings. NRA Examined; Rate-making Problems of TVA; New Deal and the Teaching of Economics; Paths of Economic Change; Business Enterprise and the Organization of Production; Changes in the

Character, Structure, and Conditions of Production; International Aspects of Problems of Production and Trade; International Movements of Capital; Our Commercial Banking System; Aspects of Co-ordination and Finance; Some Lessons Drawn from European Experience; Nationalism; Security Regulation and Speculation; Monetary Stabilization from an International Point of View; Monetary Stabilization from a National Point of View; Decentralization of Population and Industry; Co-ordination of State and Local Finance; Relief Aspects of the New Deal; Unified Program for the Unemployed. Pp. 240.

Volume XXVI, 1936

The American Economic Review, March, June, September, and December

Supplement.—Forty-eighth Annual Meeting:

Papers and Proceedings. Some Distinguishing Characteristics of the Current Recovery; Price Theories and Market Realities; Notes on Inflexible Prices; Effect of the Depression upon Earnings and Prices of Regulated and Nonregulated Industries; Size of Plants in Its Relation to Price Control and Price Flexibility; Requisites of Free Competition; Monopolistic Competition and Public Policy; Banking Act of 1935; Recent Legislation and the Banking Situation; Economic Aspects of an Integrated Social Security Program; Capital Formation; Trade Agreements Program and American Agriculture; Founding and Early History of the American Economic Association; Developments in Economic Theory; Federal Revenue Act of 1935; Relations between Federal, State, and Local Finances; Equalization of Local Government Resources; Adjustment to Instability; Transportation Problems; Fifty Years' Developments in Ideas of Human Nature and a Motivation; Institutional Economics; Place of Marginal Economics in a Collectivist System; Problem of Prices and Valuation in the Soviet System; Effects of New Deal Legislation on Industrial Relations; Report of the Fiftieth Anniversary Dinner. Pp. 350.

Supplement No. 2.—Handbook of the Association, 1936.

Volume XXVII, 1937

The American Economic Review, March, June, September, and December

Supplement.—Forty-ninth Annual Meeting:

Papers and Proceedings. Economic Interdependence, Present and Future; Quantitative and Qualitative Changes in International Trade During the Depression; Current Tendencies in Commercial Policy; Trade Problem of the Pacific; Analysis of the Nature of American Public Debts; Limits to Possible Debt Burdens, Federal, State, and Local; Debt Retirement and the Budget; United States Debt—Distribution among Holders and Present Status; Federal-State Unemployment Compensation Provisions of the Social Security Act; Unemployment Relief and Insurance; Economic Problems Arising from Social Security Taxes and Reserves; The Situation of Gold Today in Relation to World Currencies; Mechanisms and Objectives for the Control of Exchange; The Adequacy of Existing Currency Mechanisms Under Varying Circumstances; Present Situation of Inadequate Housing; Financing of Housing; Some Economic Implications of Modern Housing; Managed Currency; A Critique of Federal Personnel Policies as Applied to Professional Social Science Positions; New Opportunities for Economists and Statisticians in Federal Employment; Government Employment as a Professional Career in Economics; Indicia of Recovery; Housing and Housing Research; Distribution of Puchasing Power and Business Fluctuations; Forecast of Power Development; The Possibility of a Scientific Electrical Rate System; Co-ordination of Public and Private Power Interests in European Countries; Recent Developments in the Theory of Speculation; Control of Speculation under the Securities Exchange Act; Unorganized Speculation: the Possibility of Control. Pp. 333.

Volume XXVIII, 1938

The American Economic Review, March, June, September, and December

Supplement.—Fiftieth Annual Meeting:

Papers and Proceedings. The Significance of Marxian Economics for Present-day Economic Theory; The Significance of Marxian Economics for

Current Trends of Governmental Policy; The Rate of Interest; Security Markets and the Investment Process; Relation of Price Policy to Fluctuations of Investment; General Interest Theory; Rate of Interest; Security Regulation; Corporate Price Policies; Fiscal Policies; Rate of Consumption; Wage Rates; Social Security Program; Rate of Consumption; Durable Consumers Goods; Wage Policies. Pp. 192.

Supplement No. 2.—Handbook of the Association, 1938.

Volume XXIX, 1939

The American Economic Review, March, June, September, and December

Supplement.—Fifty-first Annual Meeting:

Papers and Proceedings. Problem of Industrial Growth in a Mature Economy; Effects of Current and Prospective Technological Developments upon Capital Formation; Public Investment in the United States; Expansion and Contraction in the American Economy; Effect of Industrial and Technological Developments upon Demand for Capital; Role of Public Investment and Consumer Capital Formation; Income and Capital Formation; Price and Production Policies of Large-Scale Enterprise; Changing Distribution Channels; Financial Control of Large-Scale Enterprise; Pure Theory of Production; Changing Character of American Industrial Relations; Wages and Hours in Relation to Innovations and Capital Formation; Effect of Wage Increase upon Employment; Relation of Wage Policies and Price Policies; An Appraisal of Factors Which Stopped Short the Recovery Development in the United States; Fiscal Policy in the Business Cycle; An Appraisal of the Workability of Compensatory Devices; Divergencies in the Development of Recovery in Various Countries; Factors Making for Change in Character of Business Cycle; Industrial Relations. Pp. 280.

Volume XXX, 1940

The American Economic Review, March, June, September, and December

Supplement.—Fifty-second Annual Meeting:

Papers and Proceedings. Objectives of Monetary Policy; Economic Issues in Social Security Policy; Bank Deposits and the Business Cycle; Problems in the Teaching of Economics; Price Control Under "Fair Trade" Legislation; Problems of American Commercial Policy; Transportation Problem; Preserving Competition Versus Regulating Monopoly; Theory of International Trade; Collective Bargaining and Job Security; Banking Reform Through Supervisory Standards; Incidence of Taxation; Economic Planning; Growth of Rigidity in Business; Economics of War; Population Problems; Cost Functions and Their Relation to Imperfect Competition. Pp. 436.

Supplement No. 2.—Handbook of the Association, 1940.

No. 5 (February, 1941)

Fifty-third Annual Meeting (December, 1940):

Papers and Proceedings. Gold and the Monetary System; Economic Research; Federal Budget; Economic Consequences of Deficit Financing; Teaching of Economics; Agricultural Situation; A Review of Fundamental Factors, an Evaluation of Public Measures, and an Appraisal of Prospects; Status and Role of Private Investment in the American Economy, 1940; Unemployment in the United States, 1930-50; Economic Consequences of War Since 1790; Some Economic Problems of War, Defense, and Postwar Reconstruction; United States in the World Economy, 1940; International Economic Relations and Problems of Commercial Policy; Price Policy and Price Behavior. Pp. 458.

Volume XXXI, 1941

The American Economic Review, March, June, September, and December

Volume XXXII, 1942

The American Economic Review, March, June, September, and December

Supplement.—Fifty-fourth Annual Meeting:

Papers and Proceedings. Economic Adjustments After Wars; Problems of Taxation; Determinants of Investment Decisions; Problems of Interna-

CALCUTTA CENTRAL LIBRARY

tional Economic Policy for the United States; History of American Corporations; Problems of Labor Market Research; Co-ordination of Federal, State, and Local Fiscal Policy; Technical Aspects of Applying a Dismissal Wage to Defense Workers; Problems of International Economic Policy; Impact of National Defense and the War upon Public Utilities; Future of Interest Rates; Effect of Managerial Policy upon the Structure of American Business; Economic Effects of Wars; Economic Aspects of Reorganization Under the Chandler Act; Economics of Industrial Research; Objectives in Applied Land Economics Curricula; Changing Position of the Banking System and Its Implications for Monetary Policy; Determination of Wages; Economic Problems of American Cities; Cost and Demand Functions of the Individual Firm; Problems of Price Control; Effects of the War and Defense Program upon Economic Conditions and Institutions; Trade Unions and the Law. Pp. 534.

Supplement No. 2.—Papers Relating to the Temporary National Economic Committee. Pp. 135.

Supplement No. 3.—Directory. Pp. 198.

Volume XXXIII, 1943

The American Economic Review, March, June, September, and December

Supplement.—Fifty-fifth Annual Meeting:

Papers and Proceedings. Economic Claims of Government and of Private Enterprise; Our Industrial Plant When Peace Comes; Financial and Government Contract Adjustments of Industry at the End of the War; Problems of Public Policy Raised by Collective Bargaining; Our Labor Force When Peace Comes; Price Control and Rationing; Case Studies in Price Control; Restoration of International Trade; Future of International Investment; International Financial Relations After the War; Economic Regionalism and Multilateral Trade; Bases of International Economic Relations; International Commodity Agreements. Pp. 508 + 15.

Volume XXXIV, 1944

The American Economic Review, March, June, September, and December

Supplement.—Fifty-sixth Annual Meeting:

Papers and Proceedings. Political Science, Political Economy, and Values; Educational Function of Economists and Political Scientists; Public Administration of Transportation under War Conditions; How Achieve Full and Stable Employment; Incentive Problems in Regulated Capitalism; Postwar Labor Problems; Social Security; Postwar Legal and Economic Position of American Women; Postwar Domestic Monetary Problems; Economic Organization of Welfare; International Trade; Regional Problems; International Monetary Problems. Pp. 440 + 16.

Supplement No. 2.—Implemental Aspects of Public Finance. Pp. 138.

Volume XXXV, 1945

The American Economic Review, March, June, September, and December

Supplement (May).—Fifty-seventh Annual Meeting:

Papers and Proceedings. Consumption Economics; Expanding Civilian Production and Employment After the War; Natural Resources and International Policy; Interdepartmental Courses in the Social Sciences; Price Control and Rationing in the War-Peace Transition; Organized Labor and the Public Interest; Aviation in the Postwar World; International Monetary and Credit Arrangements; Agricultural Price Supports and Their Consequences; Political Economy of International Cartels; Fiscal Problems of Transition and Peace; Problems of Regionalism in the United States; Food and Agriculture—Outlook and Policy; Function of Government in the Postwar American Economy. Pp. 520 + 16.

Volume XXXVI, 1946

The American Economic Review, March, June, September, and December

Supplement (May).—Fifty-eighth Annual Meeting:

Papers and Proceedings. Problem of "Full Employment"; American Economy in the Interwar Period; Postwar Labor Relations; Monetary Policy;

Changing Structure of the American Economy; Economic Problems of Foreign Areas; Publication of an Annual Review of Economics; New Frontiers in Economic Thought; Postwar Shipping Policy; Monopoly and Competition; Postwar Tax Policy; Postwar Railroad Problems; International Investment; Recent Developments in Public Utility Regulation; International Cartels; Economic Research; Methods of Focusing Economic Opinion on Questions of Public Policy (e.g., Monetary, Agricultural Price Supports); Undergraduate Teaching of Economics. Pp. 960.

Supplement No. 2.—Handbook. Pp. 143.

Volume XXXVII, 1947

The American Economic Review, March, June, September, and December

Supplement (May).—Fifty-ninth Annual Meeting:

Papers and Proceedings. Employment Act of 1946 and a System of National Bookkeeping; Social and Economic Significance of Atomic Energy; Public Debt: History, Effects on Institutions and Income, and Monetary Aspects; Economic Forecasts; Role of Social Security in a Stable Prosperity; Economic Outlook; Economy of the U.S.S.R.; Domestic versus International Economic Equilibrium; Prices: Wartime Heritage and Some Present Problems; Banking Problems; Productivity in the American Economy; International Trade Organization; Vital Problems in Labor Economics; Transportation and Public Utilities Problems; Housing Problems; Economic Research; Changing Character of Money. Pp. 781.

Volume XXXVIII, 1948

The American Economic Review, March, June, September, and December

Supplement (May).—Sixtieth Annual Meeting:

Papers and Proceedings. Economic Theory of Imperfect Competition, Oligopoly, and Monopoly; Role of Monopoly in the Colonial Trade and Expansion of Europe; Progress of Concentration in Industry; Does Large-Scale Enterprise Result in Lower Costs; Sherman Act and the Enforcement of Competition; Patent Policy; A Consideration of the Economic and Monetary Theories of J. M. Keynes; Keynesian Economics: The Propensity to Consume and the Multiplier, and Savings, Investment, and Wage Rates; Economics Collides with Ethics; An Appraisal of the Taft-Hartley Act; Fiscal Policy in Prosperity and Depression; Problems of Timing and Administering Fiscal Policy in Prosperity and Depression; Transportation and Public Utilities; Futility of Trust-Busting; National Productivity. Pp. 591.

Volume XXXIX, 1949

The American Economic Review, March, June, September, and December

Supplement (No. 1—January).—Directory. Pp. 343.

Supplement (No. 3—May).—Sixty-first Annual Meeting:

Papers and Proceedings. Commemoration of the Centenary of the Communist Manifesto—The Sociology and Economics of Class Conflict; Current Research in Business Cycles; Interregional Variations in Economic Fluctuations; Economic Research; Economic Consequences of Some Recent Antitrust Decisions; Theory and Measurement of Price Expectations; Input-Output Analysis and Its Use in Peace and War Economies; Liquidity and Uncertainty; Problems of the ITO; Commodity Marketing—Going Where; John Stuart Mill—Centennial Appraisal; Possibilities for a Realistic Theory of Entrepreneurship; Economics of Preparedness for War; Present Issues of the Latin-American Economy. Pp. 537.

Volume XL, 1950

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-second Annual Meeting:

Papers and Proceedings. What Planning and How Much Is Compatible with a Market Economy—Recent European Experience; Capitalism and Monopolistic Competition—I. The Theory of Oligopoly, II. Can the American Economy Be Made More Competitive; Capitalism and Economic

Progress; Stabilizing the Economy—The Employment Act of 1946 in Operation; Problems of an Advanced Defense Economy; Transportation in Capitalist and Socialized Economies; Can Capitalism Dispense with Free Labor Markets; Capitalism and Equality of Income; Tax Structure and Private Enterprise; Economic Power Blocs and American Capitalism; American Capitalism—Where Are We Going; U.S. Foreign Investment in Underdeveloped Areas; Economic Policy in Occupied Germany. Pp. 650.

Supplement.—Supplement to 1948 Directory. Pp. 41.

Supplement No. 2.—The Teaching of Undergraduate Economics. Pp. 226.

Volume XLI, 1951

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-third Annual Meeting:

Papers and Proceedings. Role and Interests of the Consumer; Institutional Economics; Executive Decisions at the Top Level; Economic Theory, Statistics, and Economic Practice; Capital and Interest; Economic Stabilization; Modern Industrialism and Human Values; Factors in Modern Industrial Development; Government Action or Private Enterprise in River Valley Development; Economic Progress—Some Cases, Comparisons, and Contrasts; Point Four—Development of Backward Areas; Current Problems in International Trade; Economy of the Soviet Union; Changes in National Transportation Policy; Public Utilities and National Policy; Housing Problem—Current Situation and Long-run Effects of Government Housing Programs; Economics of Medical Care—The Problem and Alternative Solutions; Economics in General Education. Pp. 816.

Supplement.—Supplement to 1948 Directory. Pp. 17.

Volume XLII, 1952

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-fourth Annual Meeting:

Papers and Proceedings. Economic Theory and Public Policy; Issues in Methodology; Business Cycle Theory; Monetary Theory; Fiscal Theory; International Trade Theory; Value Theory; Recent Developments in United States Monetary Policy; Inflation Control in the United States; American Foreign Aid Programs; International Trade in the Postwar World; Wages, Manpower, and Rearmament; Governmental Policy on Business Practices; Economic Problems of Military Mobilization; Theoretical Analysis of Economic Growth; General Factors in Economic Growth in the United States; Growth in Underdeveloped Countries; Role of War in American Economic Development; Public Utilities, Transportation, and Spatial Organization; Collective Bargaining in the Regulated Industries. Pp. 768.

Supplement.—Supplement to 1948 Directory. Pp. 11.

Volume XLIII, 1953

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-fifth Annual Meeting:

Papers and Proceedings: Monetary Policy; A Stock-taking of Bretton Woods Objectives; Underdeveloped Countries—The Theory and Practice of Technical Assistance; United States Demand for Imports; Interregional Analysis and Regional Development; Governmental Economic Activity; Development of Economic Thought; Technology; Economics in the Curricula of Agricultural Colleges in Canada and the United States; Sociology and Economics; Recent Developments in Mathematical Economics and Econometrics; Research; Long-run Effects of Full Employment on the Labor Market; Theory of Income Distribution; Distribution of Government Burdens and Benefits; Distribution and Utilization of Natural Gas. Pp. 612.

Supplement No. 2—Graduate Education in Economics. Pp. 223.

Supplement No. 3—Handbook. Pp. 187.

Volume XLIV, 1954

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-sixth Annual Meeting:

Papers and Proceedings: Fundamental Characteristics of the American Economy; An Appraisal of Economic Change; Factor Markets Versus Product Markets; Farm Prices and Farm Incomes in American Agriculture; Industrial Pricing; Technological Progress and Economic Institutions; Growth Decisions in the American Economy; Diminishing Inequality in Personal Income Distribution; Wage Determination in the American Economy; Alternative Possibilities of Inflationary Pressures and Higher Cost Bottlenecks in an Economy of Large Bargaining Units and of Less Than Pure and Perfect Competition in the Marketing of Products; Regional Wage Differentials in an Economy of Large Bargaining Units and Less Than Pure and Perfect Competition in the Marketing of Products; Automaticity of Full Employment Under the Assumption of Diminished Defense Expenditures; Institutional Aspects of Savings and Investment; Role of Corporate Taxation in the American Economy; Theory of International Trade in a World of Trade Barriers and Controls and of Variegated National Economic Systems; Economic Doctrines Implied in the Reports of the U.N. and IBRD on Underdeveloped Countries; Corporate International Investment Policies and Programs; Economic Implications of an Aging Population; Report on Graduate Training in Economics; Economics in General Education; Economic and Regulatory Problems in the Broadcasting Industry; National Transportation Policy, Pp. 765.

Supplement.—Supplement to 1953 Handbook. Pp. 11.

Volume XLV, 1955

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-seventh Annual Meeting:

Papers and Proceedings: International Flow of Economic Ideas; Development Policy in Underdeveloped Countries; Economic Development—Case Studies; Regional Economics; Urbanization and Industrialization of the Labor Force in a Developing Economy; Models of Economic Growth; Current Problems in Agricultural Economics; Long-term Trends in International Trade; Economic Research and Public Policy; Economic Stabilization, Forecasting, and the Political Process; Cyclical Experience in the Postwar Period; Debt Management and Monetary Policy; Taxation and Income Distribution; Concepts of Competition and Monopoly; Impact of Antitrust Laws; Research on the Business Firm; Price and Wage Flexibility; Pricing in Transportation and Public Utilities. Pp. 711.

Supplement.—Supplement to 1953 Handbook. Pp. 8.

Volume XLVI, 1956

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-eighth Annual Meeting:

Papers and Proceedings: Statement of the Problem of Keeping the U.S. Economy Moving Forward, But Steadily; Production and Consumption Economics of Economic Growth; Income Distribution Aspects of Expanding Production and Consumption; Increase of Consumption Part of Economic Growth; Government Expenditures and Economic Growth; Monetary Role in Balanced Economic Growth; Shortening Work Week as a Component of Economic Growth; Highway Development and Financing; Unemployment as a Phase of Economic Growth; Urban Growth and Development; Population Specter—Rapidly Declining Death Rates in Densely Populated Countries; Current Economic Thought and Its Application and Methodology in Continental Europe; Economic Thought and Its Application and Methodology in the East; Economic Potentials of Latin America; Economic Potentials of Africa; Report of the Attorney General's Committee on Antitrust Policy; Changing Patterns of Competition in Transportation and Other Utility Lines; Economics in the Curricula of Schools of Business. Pp. 651.

Volume XLVII, 1957

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-ninth Annual Meeting:

Papers and Proceedings: Institutional Economics; Economic Growth and

Development; Keynesian Economics after Twenty Years; Employment Act in the Economic Thinking of Our Times; Business Fluctuations and Inflation; Government Taxing and Spending; Developments in Banking and Monetary Policy; Price and Competitive Aspects of the Distributive Trades; Monopoly Problem as Seen by Social Scientists; Impact of Some New Developments in Economic Theory; Economics and Changing Technology; Power and Public Utility Problems; Transportation Problems; Social Security and Welfare; Income Distribution; Consumers in the American Economy; International Economics; Soviet Economic Developments; Economics in Schools. Pp. 754.

Supplement (No. 4—July).—Handbook. Pp. 522.†

† This issue is not included in the price of the volume.

Volume XLVIII, 1958

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Seventieth Annual Meeting:

Papers and Proceedings: Veblen Centenary Round Table; Democracy and Trade-Unionism; Agenda for a National Monetary Commission; Is Another Major Business Contraction Likely; Monetary Analysis and the Flow of Funds; Economic Projections and a Statistical Contribution to Price Theory; Statistical Cost Functions; Trends in Capital Investment and Capacity; Income and Consumption; State and Local Public Finance; Agricultural Parity; Further Explorations in Monopolistic-Competitive Price Theory; Petroleum and Natural Gas and the Public Interest; Critical Evaluation of Public Regulation of Independent Commissions; Current Economic Questions Relating to Western Europe; Measuring Production in the U.S.S.R.; Selected Papers—American Economic Association Competition. Pp. 677.

Supplement.—Supplement to 1965 Handbook. Pp. 32.

Volume XLIX, 1959

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Seventy-first Annual Meeting:

Papers and Proceedings: Soviet Economic Trends and Prospects; Soviet Economic Planning; Non-Russian Communist Economies; Fundamentals of Economic Progress in Underdeveloped Countries; Special Problems Facing Underdeveloped Countries; Role and Character of Foreign Aid; International Trade and Payments in an Era of Coexistence; Maintaining Full Employment and Economic Stability; Balanced Economic Growth in History—A Critique; Economics of Government Expenditures; Power Blocs and the Operation of Economic Forces; Administered Prices Reconsidered; Studies in the Classical Economics; Selected Problems in Economic Theory; Open Competition; Organization and Financing of Economic Research; Market for Economists. Pp. 689.

Volume L, 1960

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Seventy-second Annual Meeting:

Papers and Proceedings: Standards for the Performance of Our Economic System; Relations Between Economic Theory and Economic Policy; Incentives and Economic Growth: Changing Roles and Public Policies; Problems of Achieving and Maintaining: a High Rate of Economic Growth, Full Employment, and a Stable Price Level; Problem of Raising Incomes in Lagging Sectors of the Economy; Problem of International Harmony: Economic Policies for a Lasting Peace; Problem of Social Priorities; Investing in Education and Research; Facilitating Movements of Labor Out of Agriculture; Reducing Impediments to Foreign Trade; Reforming the Tax System; Improving the Efficiency of the Transportation and Utilities Systems; Research: on Theory of the Firm, on Income, Consumption, and Savings, and on Economic Development. Pp. 745.

Volume LI, 1961

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Seventy-third Annual Meeting:

Papers and Proceedings: Monetary Theory—New and Old Looks; Macroeconomic Theories of Income Distribution; Capital Theory; Managerial Economics—A New Frontier; Frontiers in Uncertainty Theory—The Evidence of Futures Markets; Distribution Costs—Concepts and Measures; Antitrust Problems; Economic Analysis of Urban Problems; Public Utilities and Transportation; Wheat—A Permanent Need for a Farm Program; Problems of Economic Instability in Other Countries; Balance of Payments of the United States—Problems and Prospects; Economics and National Security; Economic Development in Mainland China; Influence of Moral and Social Responsibility on Economic Behavior; Economic Education—Challenge to Our Profession. Pp. 675.

Volume LII, 1962

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Seventy-fourth Annual Meeting:

Papers and Proceedings: New Developments in the Theory of the Firm; Problems of Economic Development; Lagging U.S. Growth Rate; International Transmission of Business Cycles—Problems and Policies; Soviet Economic Planning; Economics of Research and Development; Studies in Business Behavior; Economic Behavior of Families; Reappraisal of the Doctrine of Consumer Sovereignty; Report of the Commission on Money and Credit; Tax Problems; Systems of Economic Accounts and Analysis for Urban Regions; Role of Transportation in Economic Development; Transportation Problems in the American Economy; Economics of Water Resource Use; Teaching of Economics; Market for Economists. Pp. 615.

Volume LIII, 1963
($8.00 a Volume)

The American Economic Review, March, June, September, and December; each 2.00

Supplement (No. 2—May).—Seventy-fifth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Public Policies with Respect to Private Business; International Commodity Stabilization; Conditions of International Monetary Equilibrium; Problems of Regional Integration; Problems of Methodology; Topics in Economic Theory; Industrial Capacity; Tax Reform; Financial Institutions and Monetary Policy—A Re-examination of Their Interrelationship; Defense and Disarmament; Pricing and Resource Allocation in Transportation and Public Utilities; Postwar Growth in the United States in the Light of the Long-swing Hypothesis; Economic Trends and Prospects in the U.S.S.R. and Eastern Europe; Japanese Economic Development; Economic Development and the Population Problem; Economic Education. Pp. 753. 4.00

Supplement (March)—Economics in the Schools. Pp. 27. .50

Volume LIV, 1964

The American Economic Review, **March, **June, **September, and **December; each 2.00

Supplement (No. 1—January).—Handbook. Pp. 464.† 8.00

**Supplement (No. 3—May).—Seventy-sixth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Theory of Monopolistic Competition after Thirty Years; Principles of Efficiency; Efficiency in Agriculture; Efficiency in the Labor Markets; Regulated Industries; Efficiency in the Use of Natural Resources; Efficiency in the Government Sector; Problems in Taxation; Financial Factors in Business Cycles; Reappraisals in American Economic History; Comparative Costs and Economic Development; Canadian-American Economic Relations—Efficiency and Independence; Efficiency of the Soviet Economy; Invited Dissertations, I and II; Efficiency in the Teaching of Economics. Pp. 698. 4.00

Supplement (March)—Surveys of Foreign Postwar Developments in Economic Thought. Pp. 55. 1.00

Supplement (June)—Surveys of Foreign Postwar Developments in Economic Thought. Pp. 101. 1.00

† This issue is not included in the price of the volume.

Volume LV, 1965

The American Economic Review, **March, June, September, and December; each	$2.00
Supplement (No. 2—May).—Seventy-seventh Annual Meeting: Papers and Proceedings: Richard T. Ely Lecture; Economic Growth—International Comparisons; Comparative Economic Systems—Nationalized Industry; Economic History—Its Contribution to Economic Education, Research, and Policy; Contribution of the History of Economic Thought to the Understanding of Economic Theory, Economic History, and the History of Economic Policy; Evolving International Monetary Mechanism; Domestic Implications of the Evolving International Monetary Mechanism; Survey Research—Three Surveys, Findings and Implications for Theory and Policy; Recent Capital and Production Theory; Technological Change—Stimuli, Constraints, Returns; New National Econometric Model—Its Application; Invited Student Dissertations; Defense Economics—Applying Economic Criteria; Nonmarket Decision Making; Economic Theory and Nonprofit Enterprise; Economics of Poverty; Economic Education—Experiments in the Teaching of Economics. Pp. 652.	4.00
Supplement (December).—The Structure of Economists' Employment and Salaries, 1964. Pp. 98.	1.00

Volume LVI, 1966
($10.00 a Volume)

The American Economic Review, March, June, September, and December; each	2.00
Supplement (No. 2—May).—Seventy-eighth Annual Meeting: Papers and Proceedings: Richard T. Ely Lecture; Allocation and Distribution Theory—Technological Innovation and Progress; Capital Theory—Technical Progress and Capital Structure; Economic Development—Advanced Technology for Poor Countries; Knowledge, Information, and Innovation in the Soviet Economy; Money and Banking—Innovations in Finance; Public Finance—Promotion of Knowledge Production and Innovation; International Economics—Progress and Transfer of Technical Knowledge; Antitrust and Patent Laws—Effects on Innovation; Public Regulation—The Impact of Changing Technology; Economics of Education; Economics of Publishing; Economics of Broadcasting and Advertising; Economics of Science Policy; Production and Use of Economic Knowledge; Labor Economics—Effects of More Knowledge. Pp. 670.	4.00
Supplement (September).—Handbook. Pp. 184.	2.50

Volume LVII, 1967

The American Economic Review, March, June, September, and December; each	2.00
Supplement (No. 2—May).—Seventy-ninth Annual Meeting: Papers and Proceedings: Richard T. Ely Lecture; Military Manpower Procurement; Cost-Benefit Analysis for Government Decisions; Economics of Health; Economic Analysis of Water Resource Problems; Transportation and Patterns of Urban Development; Antitrust and Monopoly; Economic History; Economic Development; Property Rights and Behavior; Topics in Money; Aggregate Models; Measurement of Price Change; Invited Doctoral Dissertations; *Das Kapital*—A Centenary Appreciation; Efficiency of Education in Economics. Pp. 741.	4.00

Volume LVIII, 1968

The American Economic Review, March, June, September, and December; each	2.00
Supplement (No. 2—May).—Eightieth Annual Meeting: Papers and Proceedings: Richard T. Ely Lecture; Behavioral and Ecological Economics; The Entrepreneur; Econometric Models, Their Problems and Usefulness; Regional Economic Models; Invited Student Dissertations; Problems in Pricing and Growth; Nonmarket Decision Making; Transport for Economic and Social Development; Economics of Arms Control and Disarmament; Experiments in Teaching Economics; Technological and Economic Implications of 3 Percent Growth; Economics and Noneconomics of Poverty; Economic Reform in Eastern Europe and the U.S.S.R.; International Liquidity. Pp. 751.	4.00

Supplement (June).—Surveys of National Economic Policy Issues and Policy Research. Pp. 138.	1.00
Supplement (December).—Studies of the Structure of Economists' Salaries and Income. Pp. 153.	1.00

Volume LIX, 1969

The American Economic Review, March, June, September, and December; each	2.00
Supplement (No. 2—May).—Eighty-first Annual Meeting: Papers and Proceedings: Richard T. Ely Lecture; Theory of Innovation; Research and Development and Other Determinants of Investment; Theory of the Firm and of Market Structures; Wage-Price Dynamics, Inflation, and Unemployment; Supply Function in Agriculture Revisited; Problems in the Theory of Public Choice; Efficiency of Education in Economics; Transportation and the Public Utilities; Monetary Theory; Future of Gold; Round Table on Exchange Rate Policy; Economic History; Economic Development; Problems in the Area of Poverty; Strategic Theory and Its Applications; Round Table on Allocation of Resources in Law Enforcement; Centralization and Decentralization in Economic Systems; Completed Dissertation Research. Pp. 629.	4.00

The titles of papers published in the *Papers and Proceedings*, as well as the titles of articles in the *American Economic Review*, are indexed in the *Reader's Guide* to 1953 and the *International Index* thereafter.

* * *

Address all inquiries regarding publications from 1886 through 1910 and Vols. I-LII (1911–62) of the *American Economic Review* to Walter J. Johnson, Inc., 111 Fifth Avenue, New York, New York, 10003. Inquiries regarding subsequent issues of the *Review* and other publications should be addressed to the American Economic Association office. The supply of those items double starred is exhausted.

* * *

The American Economic Association, founded, among other purposes, for the "encouragement of economic research" and "the encouragement of perfect freedom of economic discussion," has over seventeen thousand members, including public and professional men and most of the leading students of political economy in America. Membership dues are ten dollars a year. Each member receives all current reports and publications of the Association, including the *Journal of Economic Literature*. In addition to membes, there are some seven thousand subscribers to the *American Economic Review*, institutional and individual. Applications for membership and other inquiries should be sent to:

Harold F. Williamson
SECRETARY OF THE AMERICAN ECONOMIC ASSOCIATION
629 Noyes Street, Evanston, Illinois 60201

New and Notable 1817 *Harper Texts*

Lipsey and Steiner's
ECONOMICS
Second Edition

Changes in the Second Edition: The structure is greatly modified. The macroeconomics segment has been completely rewritten, reorganized, and more than doubled in size. Overall, the text is somewhat longer, although there are now 46 chapters instead of 58. Throughout, scores of newly added examples and tables show up-to-date, real world data. There is a *much greater emphasis on policy material,* with the policy sections carefully integrated with the rest of the text. *845 pp.; $9.95*

Available adjuncts:

- **Instructor's Manual**
- **New Study Guide** (*326 pp.; $3.50*)
- **Answers to Problems and Cases in the Study Guide**

THE ECONOMICS OF MONEY AND BANKING, Fifth Edition
LESTER V. CHANDLER, *Princeton University*
576 pp.; $9.95

ECONOMIC STATISTICS AND ECONOMETRICS
An Introduction to Quantitative Economics
EDWARD J. KANE, *Boston College*
437 pp.; $11.95

MACROECONOMIC ACTIVITY
Theory, Forecasting, and Control—An Econometric Approach
MICHAEL K. EVANS, *University of Pennsylvania*
627 pp.; $15.95

STATISTICS, AN INTRODUCTORY ANALYSIS
Second Edition
TARO YAMANE, *New York University*
919 pp.; $11.50

MATHEMATICAL ANALYSIS
Business and Economic Applications
JEAN E. DRAPER, *University of Arizona,* and
JANE S. KLINGMAN, *Xavier University, Ohio*
568 pp.; $12.50

Harper & Row, Publishers, 49 East 33d Street, N.Y. 10016

Significant reading from Basic Books

THE ECONOMIC THEORY OF "MANAGERIAL" CAPITALISM

by Robin Marris, King's College, Cambridge

While studies have long established that today's top manager is not primarily interested in profits, Mr. Marris is the first writer to demonstrate that profit maximization is not a *theoretically logical* objective for the large diversified corporation, whereas growth is.

"In this brilliant book, Mr. Marris is laying siege to one of the most fundamental assumptions of economics. If the siege is successful, as I am persuaded it must be, it will have a particularly grave unhinging effect."—*from the Foreword by John Kenneth Galbraith*

Now available in paperback ***$3.95***

CHILDREN, POVERTY, AND FAMILY ALLOWANCES

by James C. Vadakin, University of Miami

With a Foreword by Daniel P. Moynihan

"... the best book ever written on family allowances."—*New Republic* ***$6.50***

New perspectives on poverty

These two volumes "represent the best professional thinking on this crucial subject and suggest some new, promising guidelines."—*Publishers' Weekly*

ON UNDERSTANDING POVERTY

Perspectives from the Social Sciences

Edited by Daniel P. Moynihan, Assistant to the President on urban affairs

Daniel P. Moynihan, Peter Rossi, S. M. Miller, Herbert J. Gans, Oscar Lewis, and other leading social scientists sort out the various conceptions of poverty now abroad and try to reach some judgment about the evidence in support of them.

To be published June, 1969 ***$10.00***

ON FIGHTING POVERTY

Perspectives from Experience

Edited by James L. Sundquist, The Brookings Institution

James Sundquist, Adam Yarmolinsky, Sanford Kravitz, Robert Coles, and other veterans of the war on poverty describe how basic strategic and technical decisions were made and attempt to evaluate our experience thus far.

To be published June, 1969 ***$8.50***

At your bookseller, or order from

BASIC BOOKS, INC.
404 Park Ave. South, New York, N.Y. 10016

Recently Published!

For Comparative Economics courses and courses on the Soviet Economy

THE SOVIET ECONOMY
Howard J. Sherman
University of California, Riverside

This new and comprehensive text begins with an introduction to the basic institutions and ideology of the Soviet economy. It traces the evolution of the Soviet economy since 1917 and deals at length with the vital question of the functioning of central planning in theory and in practice.

An especially interesting feature is the examination of recent controversies over reform and decentralization in the Soviet Union and Eastern Europe. Careful comparisons with the performance of the United States' economy are presented each step of the way.

Paperbound
371 pages
$4.50

Little, Brown and Company
34 Beacon Street
Boston, Massachusetts 02106

nothing old much, that's new...

THE AMERICAN ECONOMY: Analysis and Policy

Sanford D. Gordon, George G. Dawson, and Jess Witchel

The book is designed for a principles course in economics. More emphasis is placed on the analytical tools of economics than on the descriptive approach. In easily understood language the authors carefully explain all the major economic concepts found in the most rigorous books, but with far less mathematics and in less abstract form, with examples relevant to the average college student. Standardized tests have shown that average college students perform better when using the approach and materials developed in this text than other well-known texts. This book also prepares the student for advanced courses because it covers the major concepts.

Color inserts use graphs, explanatory text, and transparent overlays for visual summaries of the classical and the Keynesian economic models.

A Visual Analysis of the American Economy, prepared by Sanford D. Gordon, consists of overhead projector transparencies (54 acetates, presenting 18 concepts), provided for use with this text, other texts, or alone. An Instructor's Manual accompanies **The American Economy.**

The American Economy: 1969 495 pages $8.50

A Visual Analysis of the American Economy: $64.50

AMERICAN AGRICULTURE: The Changing Structure — Wyn F. Owen

This volume focuses on the major changes taking place in the modern agricultural economy: the trend toward large farming units, the relation of farming units to agriculturally based non-farm businesses, the interrelationship of these non-farm businesses, the structure of agricultural finance and the integration of the farm economy into the national economy. The need for revised public policy is clearly indicated.

1969 Approx. 160 pages Paperbound $1.95

RAYTHEON D. C. HEATH AND COMPANY
A DIVISION OF RAYTHEON EDUCATION COMPANY
285 Columbus Avenue, Boston, Massachusetts 02116
In Canada: D. C. Heath, Canada: A Division of Raytheon Canada, Limited
Suite 1408, 100 Adelaide Street West, Toronto 1, Ontario

NEW BOOKS FROM HOUGHTON MIFFLIN

ECONOMIC ESSENTIALS: A Core Approach
Roger C. Van Tassel, Clark University
400 pages, Paper, April 1969. **An Instructor's Manual is available.**

ADMINISTRATIVE INTELLIGENCE-INFORMATION SYSTEMS
Preston P. LeBreton, University of Washington
308 pages, 1969, Paper, $4.25.

PRICE THEORY AND ITS USES—Second Edition
Donald Stevenson Watson, George Washington University
443 pages, 1968, $7.50

PRICE THEORY IN ACTION: A Book of Reading—Second Edition
Donald Stevenson Watson, George Washington University
400 pages, 1969, Paper, $4.75.

COST ACCOUNTING: Principles and Managerial Applications Second Edition
Gerald Crowningshield, Rider College
812 pages, 1969, $9.95. **An Instructor's Manual is available.**

READINGS IN INTERNATIONAL ACCOUNTING
Edited by **Kenneth B. Berg, Gerhard G. Mueller,** and **Lauren M. Walker,** all of University of Washington
305 pages, 1969, Paper, $4.75.

MATHEMATICS OF FINANCE—Third Edition
Robert and Helen Cissell, Xavier University, Ohio
346 pages, 1969, $7.50. **An Instructor's Manual is available.**

INTERNATIONAL MARKETING MANAGEMENT
Readings in Systems and Method
Edited by **Michael J. Thomas,** Syracuse University
496 pages, 1969, Paper, $5.95.

DISTRIBUTION CHANNELS: BEHAVIORAL DIMENSIONS
Louis W. Stern, The Ohio State University
305 pages, 1969, Paper, $4.25.

Houghton Mifflin *Boston / Atlanta / Dallas / Geneva, Ill. / New York / Palo Alto*

Third Edition

ECONOMICS: A GENERAL INTRODUCTION

By Lloyd G. Reynolds, *Yale University*

This Third Edition of one of the leading texts for introductory courses in economics has been so extensively revised that it could be viewed as a completely new volume. One third of the book is new writing, half of the chapters have been heavily rewritten, and three new chapters have been added, while at the same time retaining those features that made the previous texts so popular and teachable.

Emerging policy problems have been given greater emphasis in the areas of: poverty, economics of education and manpower development, economics of government production, including cost-benefit analysis, federal-state-local fiscal relations, reconversion problems after Vietnam, the "run on gold" and the new two-price system.

New ideas and research findings have been worked into the text where appropriate to make it intellectually, as well as factually up to date. Careful retouching of all theoretical sections has resulted in better sequencing, greater clarity, and fuller definitions of basic concepts. There is a shift to macro-micro sequence, though it is flexible enough to be equally teachable in a micro-macro course. As before, Reynold's *Economics* thoroughly covers the basic principles of economics in a logical, concise presentation that avoids unnecessary wordage.

852 pages/Text price $9.50

THE REVIEW GUIDE AND WORKBOOK by John J. Arena and Richard C. Porter has also been completely revised and is more fully integrated with the text. (Text price $3.95) Again, an INSTRUCTOR'S MANUAL by Professor Reynolds will be available to adopters.

Coming in August

CONTEMPORARY ECONOMIC ISSUES

Edited by Neil W. Chamberlain

Six original articles written especially for this volume on current issues of urban problems, poverty, health services, higher education, transportation, and population. Authors are: Wilbur Thompson, Mary Jean Bowman, Herman W. Somers, Allan Cartter, William Vickrey, and Richard Easterlin.

PAPERBOUND

Write for Examination Copies

RICHARD D. IRWIN, INC. — Homewood, Illinois 60430

Oxford University Press

On Keynesian Economics and the Economics of Keynes:

A STUDY IN MONETARY THEORY

By AXEL LEIJONHUFVUD, *University of California, Los Angeles.* "This is a tremendous image shattering book. It certainly opens up new vistas of what Keynes really meant! It will become a classic in its own right."—James M. Rock, University of Utah

1968 420 pp. $7.00

Economic Accounting and Development Planning

By BRIAN VAN ARKADIE, *Ministry of Economic Affairs and Development Planning, Dar es Salaam, and* CHARLES R. FRANK, JR., *Princeton University*

1966 406 pp. tables paper $4.95

(Revised American Edition to be published Fall, 1969, cloth $7.50)

Critical Essays in Monetary Theory

By SIR JOHN HICKS, *Oxford University*

1967 256 pp. $6.00

Leading Issues in Development Economics:

SELECTED MATERIALS AND COMMENTARY

By GERALD M. MEIER, *Stanford University*

1964 592 pp. $7.50

Modern Capitalism:

THE CHANGING BALANCE OF PUBLIC AND PRIVATE POWER

By ANDREW SHONFIELD, *Royal Institute of International Affairs*

1965 475 pp. cloth $10.50 paper $4.50

Essays in Economics:

THEORIES AND THEORIZING

By WASSILY LEONTIEF, *Harvard University*

1966 272 pp. cloth $6.00 paper $2.50

OXFORD UNIVERSITY PRESS
200 Madison Avenue, New York, N.Y. 10016

NOTICE

Members of the American Economic Association
and
Subscribers to the *American Economic Review*

CHANGE IN NAME

The name of the *Journal of Economic Abstracts* has been changed to the *Journal of Economic Literature,* effective with the March, 1969, issue. The new *Journal* has been expanded to contain original articles and book reviews and bibliographical listings formerly in the *American Economic Review* and will contain the material that has been in the *Journal of Economic Abstracts.*

Members of the Association will receive both the *American Economic Review* and the *Journal of Economic Literature.* Subscribers to the *Review* will receive the *Journal* without extra charge, as has been true with the *Journal of Economic Abstracts* since 1966.

Please do not place a separate order for the *Journal of Economic Literature* unless you want an extra copy ($6.00 a year). The separate order should be sent with check to: Universal Mailers, 317 South Wacker Drive, Chicago, Illinois, 60606.

American Economic Association
629 Noyes Street
Evanston, Illinois 60201